Important Probability Distributions

(Discrete and Continuous)

<u>Lecture-02</u>

Dr. Tahseen A. Jilani
Postdoc (England-UK), PhD (UoK, HEC Indigenous program),
MSc(Statistics), MA(Economics), BSc(CS)

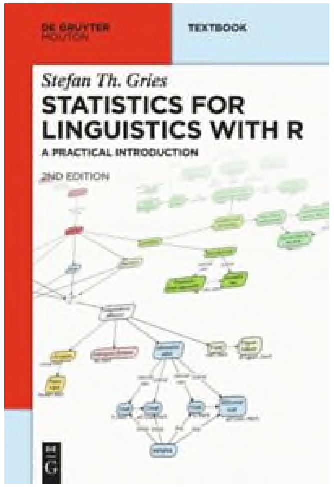
Associate Professor, Department of Computer Science, University of Karachi

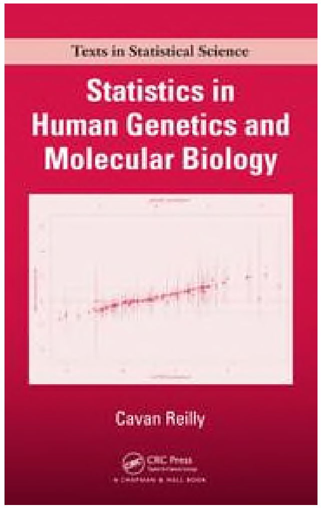
Overview

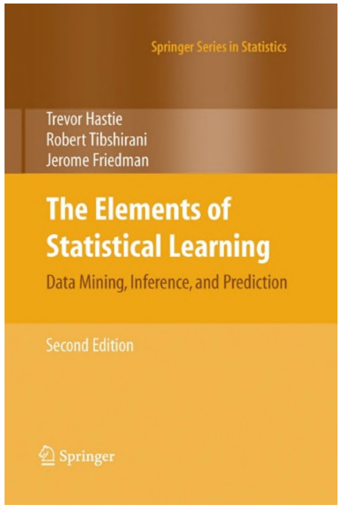
- Types of probability distributions.
- Bernoulli and Binomial probability distribution
- Poisson probability distribution
- Binomial probability approximation using Poisson probability distribution

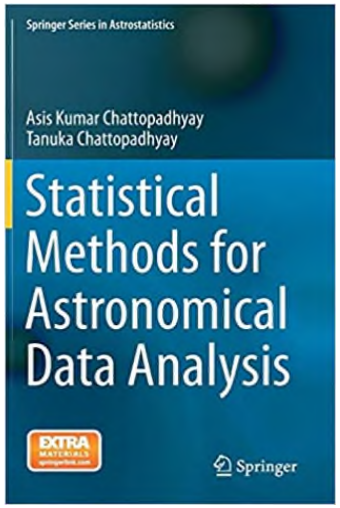
Some common probability distributions

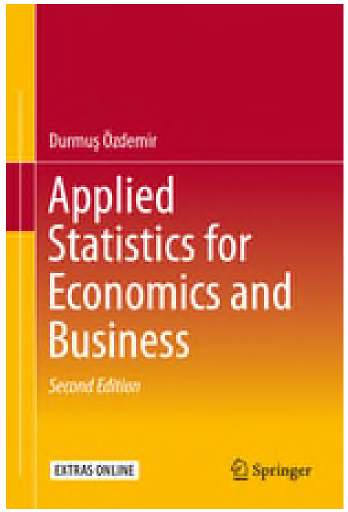
- There are hundreds of probability distributions in various fields. Such as engineering, genetics, social sciences, linguistics, business and economics etc.
- We will cover few most common probability distributions along with their assumptions, properties and mathematics moments.











Some Important probability distributions

- Discrete Probability distributions
  - Bernoulli and Binomial distributions
  - Poisson distribution
  - Geometric and Negative binomial distributions.

- Continuous Probability distributions
  - Uniform distribution
  - Exponential and Gamma distributions
  - Normal distribution

- Asymptotic analyses:
  - Normal approximation to Binomial distribution.
  - Normal approximation to Poisson distribution.

- Note: here distribution means probability distribution

Bernoulli and Binomial Probability Distributions

Bernoulli Probability Distribution

- Only one trial of the experiment.
- There are only two possible events/ outcomes success (S) or failure (F).
- P(Success) = p and P(Failure) = q =1- p
- The two events are mutually exclusive (either success occur or failure occurs but not both together).
- P(S) + P(F) = 1 = P(S)
- Examples
  - One coin tossing where P(H) +P(T) =1.
  - D.o.S attack on a computer P(D.o.S Attack) + P(Not a D.o.S attack) =1.
  - Patient diseased or not-diseased.

Bernoulli Probability Distribution

- Let X be the random variable showing success in the experiment. Then the mathematical form of the Bernoulli distribution is:

$$P(X = x) = p^x\, q^{1-x} \text{ where } x = 0, 1$$

X=x	0	1
P(X=x)	q =1 - p	p
x.P(x)	0	p
x^2.P(x)	0	p

Mean = E(X) = $\sum_{x=0}^{1} xP(x)$ = 0. q + 1.p = p

E(X^2) = $\sum_{x=0}^{1} x^2 P(x)$ = 0^2. q + 1^2.p = p

Therefore **Variance**= E(X^2) – $[E(x)]^2$ = p – p^2 = p (1-p) = pq

Binomial Probability Distribution

- An extension of Bernoulli distribution with n independent Bernoulli trials (if n=1 its Bernoulli random variable).
  - e.g., 15 tosses of a coin; 20 patients; 1000 people surveyed.
- A binary outcome
  - e.g., head or tail in each toss of a coin; disease or no disease
  - Generally called "success" and "failure"
  - P(S) + P(F) = 1 = P(S)
- Constant probability for each observation
  - e.g., Probability of getting a tail is the same each time we toss the coin.
  - e.g. Probability(survival) is the same for each patient. *(strange…)*
- The sequence order of success or failures not important.
  - e.g. in n=3 trials of a coin HHT, HTH and THH have same meaning in binomial experiments.

FOCUS - Binomial distribution -Example

- Example: In three coins tossing (or tossing a coin three times).
  - $n = 3$ with $N(S) = 2^3 = 8$;
  - With S = {HHH, <u>HHT,HTH,THH</u>, <u>THT, TTH, HTT</u>, TTT}
  - Three success in n=3 trials: P(three heads) $= {}^3C_3$ P(H). P(H). P(H)
  - Two successes in n=3 trials: P(Two heads) $= {}^3C_2$ P(H). P(H). P(T)
  - Only one success in n=3 trial: P(One heads) $= {}^3C_1$ P(H). P(T). P(T)
  - No success in n = 3 trials: P(three heads) $= {}^3C_0$ P(T).P(T).P(T)

- We can write in probability format as:
  - $P(\text{three heads}) = P(X=3) = {}^3C_3\, p.p.p = {}^3C_3\, p^3\, q^{3-3} = {}^3C_3\, p^3\, q^0$
  - $P(\text{two heads}) = P(X=2) = {}^3C_2\, p.p.q = {}^3C_2\, p^2\, q^{3-2} = {}^3C_2\, p^2\, q^1$
  - $P(\text{One heads}) = P(X=1) = {}^3C_1\, p.q.q = {}^3C_1\, p\, q^{3-1} = {}^3C_1\, p\, q^2$
  - $P(\text{No heads}) = P(X=0) = {}^3C_0\, q.q.q = {}^3C_0\, p^0\, q^{3-0} = {}^3C_0\, p^0\, q^3$

Binomial probability distribution

- The mathematical formula for Binomial distribution is:

$$P(X = x) = C_x^n \, p^x \, q^{n-x} \text{ where } x = 0,1,2,\ldots,n$$

- Here X is the random variable showing number of success in n independent trials. ***X ~ Bin (n, p)***
- C_x^n shows the number of possible ways to get x successes in n trials.
- P = P(Success) and q= P(Failure) = 1-p.
- Minimum x=0 (No success) ; x=n (all trials give success)
- Mean = np and variance = npq (quite similar to that of Bernoulli pd).

Binomial distribution - Examples

1. If I toss a coin 20 times, the probability of getting exactly 10 heads?

$$\binom{20}{10}(.5)^{10}(.5)^{10} = .176$$

2. If I toss a coin 20 times, what's the probability of getting of getting 2 or fewer heads? $P(X \leq 2) = ?$

$$\binom{20}{0}(.5)^{0}(.5)^{20} = \frac{20!}{20!0!}(.5)^{20} = 9.5x10^{-7} +$$

$$\binom{20}{1}(.5)^{1}(.5)^{19} = \frac{20!}{19!1!}(.5)^{20} = 20x9.5x10^{-7} = 1.9x10^{-5} +$$

$$\binom{20}{2}(.5)^{2}(.5)^{18} = \frac{20!}{18!2!}(.5)^{20} = 190x9.5x10^{-7} = 1.8x10^{-4}$$

$$= 1.8x10^{-4}$$

Your turn: Let X be the random variable showing number of heads tossed in 5 coin tosses.

Binomial distribution table b(x,n,p)

Numbers in the table represent $p(X=x)$ for a binomial distribution with n trials and probability of success p.

		p										
n	*x*	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.9
1	0	0.900	0.800	0.750	0.700	0.600	0.500	0.400	0.300	0.250	0.200	0.100
	1	0.100	0.200	0.250	0.300	0.400	0.500	0.600	0.700	0.750	0.800	0.900
2	0	0.810	0.640	0.563	0.490	0.360	0.250	0.160	0.090	0.063	0.040	0.010
	1	0.180	0.320	0.375	0.420	0.480	0.500	0.480	0.420	0.375	0.320	0.180
	2	0.010	0.040	0.063	0.090	0.160	0.250	0.360	0.490	0.563	0.640	0.810
3	0	0.729	0.512	0.422	0.343	0.216	0.125	0.064	0.027	0.016	0.008	0.001
	1	0.243	0.384	0.422	0.441	0.432	0.375	0.288	0.189	0.141	0.096	0.027
	2	0.027	0.096	0.141	0.189	0.288	0.375	0.432	0.441	0.422	0.384	0.243
	3	0.001	0.008	0.016	0.027	0.064	0.125	0.216	0.343	0.422	0.512	0.729
4	0	0.656	0.410	0.316	0.240	0.130	0.063	0.026	0.008	0.004	0.002	0.000
	1	0.292	0.410	0.422	0.412	0.346	0.250	0.154	0.076	0.047	0.026	0.004
	2	0.049	0.154	0.211	0.265	0.346	0.375	0.346	0.265	0.211	0.154	0.049
	3	0.004	0.026	0.047	0.076	0.154	0.250	0.346	0.412	0.422	0.410	0.292
	4	0.000	0.002	0.004	0.008	0.026	0.063	0.130	0.240	0.316	0.410	0.656
5	0	0.590	0.328	0.237	0.168	0.078	0.031	0.010	0.002	0.001	0.000	0.000
	1	0.328	0.410	0.396	0.360	0.259	0.156	0.077	0.028	0.015	0.006	0.000
	2	0.073	0.205	0.264	0.309	0.346	0.313	0.230	0.132	0.088	0.051	0.008
	3	0.008	0.051	0.088	0.132	0.230	0.313	0.346	0.309	0.264	0.205	0.073
	4	0.000	0.006	0.015	0.028	0.077	0.156	0.259	0.360	0.396	0.410	0.328
	5	0.000	0.000	0.001	0.002	0.010	0.031	0.078	0.168	0.237	0.328	0.590
6	0	0.531	0.262	0.178	0.118	0.047	0.016	0.004	0.001	0.000	0.000	0.000
	1	0.354	0.393	0.356	0.303	0.187	0.094	0.037	0.010	0.004	0.002	0.000
	2	0.098	0.246	0.297	0.324	0.311	0.234	0.138	0.060	0.033	0.015	0.001
	3	0.015	0.082	0.132	0.185	0.276	0.313	0.276	0.185	0.132	0.082	0.015
	4	0.001	0.015	0.033	0.060	0.138	0.234	0.311	0.324	0.297	0.246	0.098
	5	0.000	0.002	0.004	0.010	0.037	0.094	0.187	0.303	0.356	0.393	0.354
	6	0.000	0.000	0.000	0.001	0.004	0.016	0.047	0.118	0.178	0.262	0.531
7	0	0.478	0.210	0.133	0.082	0.028	0.008	0.002	0.000	0.000	0.000	0.000
	1	0.372	0.367	0.311	0.247	0.131	0.055	0.017	0.004	0.001	0.000	0.000
	2	0.124	0.275	0.311	0.318	0.261	0.164	0.077	0.025	0.012	0.004	0.000
	3	0.023	0.115	0.173	0.227	0.290	0.273	0.194	0.097	0.058	0.029	0.003
	4	0.003	0.029	0.058	0.097	0.194	0.273	0.290	0.227	0.173	0.115	0.023
	5	0.000	0.004	0.012	0.025	0.077	0.164	0.261	0.318	0.311	0.275	0.124
	6	0.000	0.000	0.001	0.004	0.017	0.055	0.131	0.247	0.311	0.367	0.372
	7	0.000	0.000	0.000	0.000	0.002	0.008	0.028	0.082	0.133	0.210	0.478

(continued)

Binomial distribution table b(x,n,p)

Numbers in the table represent $p(X=x)$ for a binomial distribution with n trials and probability of success p.

		p										
n	x	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.9
8	0	0.430	0.168	0.100	0.058	0.017	0.004	0.001	0.000	0.000	0.000	0.000
	1	0.383	0.336	0.267	0.198	0.090	0.031	0.008	0.001	0.000	0.000	0.000
	2	0.149	0.294	0.311	0.296	0.209	0.109	0.041	0.010	0.004	0.001	0.000
	3	0.033	0.147	0.208	0.254	0.279	0.219	0.124	0.047	0.023	0.009	0.000
	4	0.005	0.046	0.087	0.136	0.232	0.273	0.232	0.136	0.087	0.046	0.005
	5	0.000	0.009	0.023	0.047	0.124	0.219	0.279	0.254	0.208	0.147	0.033
	6	0.000	0.001	0.004	0.010	0.041	0.109	0.209	0.296	0.311	0.294	0.149
	7	0.000	0.000	0.000	0.001	0.008	0.031	0.090	0.198	0.267	0.336	0.383
	8	0.000	0.000	0.000	0.000	0.001	0.004	0.017	0.058	0.100	0.168	0.430
9	0	0.387	0.134	0.075	0.040	0.010	0.002	0.000	0.000	0.000	0.000	0.000
	1	0.387	0.302	0.225	0.156	0.060	0.018	0.004	0.000	0.000	0.000	0.000
	2	0.172	0.302	0.300	0.267	0.161	0.070	0.021	0.004	0.001	0.000	0.000
	3	0.045	0.176	0.234	0.267	0.251	0.164	0.074	0.021	0.009	0.003	0.000
	4	0.007	0.066	0.117	0.172	0.251	0.246	0.167	0.074	0.039	0.017	0.001
	5	0.001	0.017	0.039	0.074	0.167	0.246	0.251	0.172	0.117	0.066	0.007
	6	0.000	0.003	0.009	0.021	0.074	0.164	0.251	0.267	0.234	0.176	0.045
	7	0.000	0.000	0.001	0.004	0.021	0.070	0.161	0.267	0.300	0.302	0.172
	8	0.000	0.000	0.000	0.000	0.004	0.018	0.060	0.156	0.225	0.302	0.387
	9	0.000	0.000	0.000	0.000	0.000	0.002	0.010	0,040	0.075	0.134	0.387
10	0	0.349	0.107	0.056	0.028	0.006	0.001	0.000	0.000	0.000	0.000	0.000
	1	0.387	0.268	0.188	0.121	0.040	0.010	0.002	0.000	0.000	0.000	0.000
	2	0.194	0.302	0.282	0.233	0.121	0.044	0.011	0.001	0.000	0.000	0.000
	3	0.057	0.201	0.250	0.267	0.215	0.117	0.042	0.009	0.003	0.001	0.000
	4	0.011	0.088	0.146	0.200	0.251	0.205	0.111	0.037	0.016	0.006	0.000
	5	0.001	0.026	0.058	0.103	0.201	0.246	0.201	0.103	0.058	0.026	0.001
	6	0.000	0.006	0.016	0.037	0.111	0.205	0.251	0.200	0.146	0.088	0.011
	7	0.000	0.001	0.003	0.009	0.042	0.117	0.215	0.267	0.250	0.201	0.057
	8	0.000	0.000	0.000	0.001	0.011	0.044	0.121	0.233	0.282	0.302	0.194
	9	0.000	0.000	0.000	0.000	0.002	0.010	0.040	0.121	0.188	0.268	0.387
	10	0.000	0.000	0.000	0.000	0.000	0.001	0.006	0.028	0.056	0.107	0.349
11	0	0.314	0.086	0.042	0.020	0.004	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.384	0.236	0.155	0.093	0.027	0.005	0.001	0.000	0.000	0.000	0.000
	2	0.213	0.295	0.258	0.200	0.089	0.027	0.005	0.001	0.000	0.000	0.000
	3	0.071	0.221	0.258	0.257	0.177	0.081	0.023	0.004	0.001	0.000	0.000
	4	0.016	0.111	0.172	0.220	0.236	0.161	0.070	0.017	0.006	0.002	0.000
	5	0.002	0.039	0.080	0.132	0.221	0.226	0.147	0.057	0.027	0.010	0.000
	6	0.000	0.010	0.027	0.057	0.147	0.226	0.221	0.132	0.080	0.039	0.002
	7	0.000	0.002	0.006	0.017	0.070	0.161	0.236	0.220	0.172	0.111	0.016
	8	0.000	0.000	0.001	0.004	0.023	0.081	0.177	0.257	0.258	0.221	0.071
	9	0.000	0.000	0.000	0.001	0.005	0.027	0.089	0.200	0.258	0.295	0.213
	10	0.000	0.000	0.000	0.000	0.001	0.005	0.027	0.093	0.155	0.236	0.384
	11	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.020	0.042	0.086	0.314

Binomial distribution table b(x,n,p)

Numbers in the table represent $p(X=x)$ for a binomial distribution with n trials and probability of success p.

		p										
n	*x*	0.1	0.2	0.25	0.3	0.4	0.5	0.6	0.7	0.75	0.8	0.9
15	*0*	0.206	0.035	0.013	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.343	0.132	0.067	0.031	0.005	0.000	0.000	0.000	0.000	0.000	0.000
	2	0.267	0.231	0.156	0.092	0.022	0.003	0.000	0.000	0.000	0.000	0.000
	3	0.129	0.250	0.225	0.170	0.063	0.014	0.002	0.000	0.000	0.000	0.000
	4	0.043	0.188	0.225	0.219	0.127	0.042	0.007	0.001	0.000	0.000	0.000
	5	0.010	0.103	0.165	0.206	0.186	0.092	0.024	0.003	0.001	0.000	0.000
	6	0.002	0.043	0.092	0.147	0.207	0.153	0.061	0.012	0.003	0.001	0.000
	7	0.000	0.014	0.039	0.081	0.177	0.196	0.118	0.035	0.013	0.003	0.000
	8	0.000	0.003	0.013	0.035	0.118	0.196	0.177	0.081	0.039	0.014	0.000
	9	0.000	0.001	0.003	0.012	0.061	0.153	0.207	0.147	0.092	0.043	0.002
	10	0.000	0.000	0.001	0.003	0.024	0.092	0.186	0.206	0.165	0.103	0.010
	11	0.000	0.000	0.000	0.001	0.007	0.042	0.127	0.219	0.225	0.188	0.043
	12	0.000	0.000	0.000	0.000	0.002	0.014	0.063	0.170	0.225	0.250	0.129
	13	0.000	0.000	0.000	0.000	0.000	0.003	0.022	0.092	0.156	0.231	0.267
	14	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.031	0.067	0.132	0.343
	15	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.013	0.035	0.206
20	*0*	0.122	0.012	0.003	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	1	0.270	0.058	0.021	0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	2	0.285	0.137	0.067	0.028	0.003	0.000	0.000	0.000	0.000	0.000	0.000
	3	0.190	0.205	0.134	0.072	0.012	0.001	0.000	0.000	0.000	0.000	0.000
	4	0.090	0.218	0.190	0.130	0.035	0.005	0.000	0.000	0.000	0.000	0.000
	5	0.032	0.175	0.202	0.179	0.075	0.015	0.001	0.000	0.000	0.000	0.000
	6	0.009	0.109	0.169	0.192	0.124	0.037	0.005	0.000	0.000	0.000	0.000
	7	0.002	0.055	0.112	0.164	0.166	0.074	0.015	0.001	0.000	0.000	0.000
	8	0.000	0.022	0.061	0.114	0.180	0.120	0.035	0.004	0.001	0.000	0.000
	9	0.000	0.007	0.027	0.065	0.160	0.160	0.071	0.012	0.003	0.000	0.000
	10	0.000	0.002	0.010	0.031	0.117	0.176	0.117	0.031	0.010	0.002	0.000
	11	0.000	0.000	0.003	0.012	0.071	0.160	0.160	0.065	0.027	0.007	0.000
	12	0.000	0.000	0.001	0.004	0.035	0.120	0.180	0.114	0.061	0.022	0.000
	13	0.000	0.000	0.000	0.001	0.015	0.074	0.166	0.164	0.112	0.055	0.002
	14	0.000	0.000	0.000	0.000	0.005	0.037	0.124	0.192	0.169	0.109	0.009
	15	0.000	0.000	0.000	0.000	0.001	0.015	0.075	0.179	0.202	0.175	0.032
	16	0.000	0.000	0.000	0.000	0.000	0.005	0.035	0.130	0.190	0.218	0.090
	17	0.000	0.000	0.000	0.000	0.000	0.001	0.012	0.072	0.134	0.205	0.190
	18	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.028	0.067	0.137	0.285
	19	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.007	0.021	0.058	0.270
	20	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.003	0.012	0.122

Binomial distribution - Examples

- The probability that a patient recovers from a rare blood disease is 0.4. If 15 people are known to have contacted this disease, what is the probability that:
  - At least 10 survive?
  - From 3 to 8 survive?
  - Exactly 5 survive?
  - At most 12 survive?
- Let the random variable represents the number of people that survive. X = 0,1,2,…,15. where n=15 and P(survive)= 0. = p.

1. $P(X \geq 10) = 1 - P(X < 10) = 1 - P(X \leq 9)$

$$= 1 - \sum_{x=0}^{9} b(x; 15, 0.4) = 1 - 0.9662 = 0.0338$$

Binomial distribution - Examples

From 3 to 8 survive?

$P(3 \ \leq X \leq 8) = P(X \leq 8) - P(X < 3) = P(X \leq 8) - P(X \leq 2)$

$= \sum_{x=0}^{8} b(x; 15{,}0.4) \ - \sum_{x=0}^{2} b(x; 15{,}0.4) = 0.9050 - 0.0271.$

$P(3 \ \leq X \leq 8) = 0.8779$

From 3 to 8 survive?

$$P(X = 5) = b(x = 5; 15{,}0.4) = \sum_{x=0}^{5} b(x; 15{,}0.4) \ - \sum_{x=0}^{4} b(x; 15{,}0.4)$$

$= 0.4032 - 0.2173 = 0.1859$

At most 12 survive?

$P(X \leq 12) = \sum_{x=0}^{12} b(x; 15{,}0.4) =$

Binomial distribution - Examples

- The mean and variance for above example are:

  - Mean = E(X) = np = 15 x 0.4 = 6

  (On average 6 people will survive)

  - Variance = V(X) = npq = 15 x 0.4 x 0.6 = 3.6

  (And so standard deviation = 1.8974). The variation in mean number of survived people could be 1.8974.

Practice Problem: Binomial distribution

Q. You are conducting a case-control study of smoking and lung cancer. If the probability of being a smoker among lung cancer cases is 0.6, what's the probability that in a group of 8 cases you have:

(Use Binomial Table)

a. Less than 2 smokers? $P(X < 2) = P(X \leq 1) = ?$

b. More than 5? $P(X>5) = P(X \geq 6) = 1 - P(X \leq 5) = ?$

c. What are the expected value and variance of the number of smokers?

$$\text{Mean} = E(X) = n.p = 8 \times 0.6 = 4.8$$

$$\text{Variance} = V(X) = n.p.q = 8 \times 0.6 \times 0.4 = 1.92$$

Questions to do (Mental maths type questions)

- A fair, 4-sided die has the numbers 1, 2, 3, 4 on its faces. The die is rolled 20 times. The random variable X represents the number of 4s obtained.
  a) Find the mean and variance of X.
  b) Find $P(X < \mu - \sigma)$. Here μ is the mean (np) and σ is the standard deviation ($\sqrt{npq}$)

- David believes that 35% of people in a certain town will vote for him in the next election and he commissions a survey. Find the minimum number of people the survey should ask to have a mean number of more than 100 voting for David.

- An examiner is trying to design a multiple choice test. For students answering the test at random, he requires that the mean score on the test should be 20 and standard deviation at least 4. Find how many choices each question should have and the number of questions there should be. Number of choices is fixed across questions and show be as small as possible.

Answer

2. What's the probability that **at most** 10 exposed subjects develop the disease?

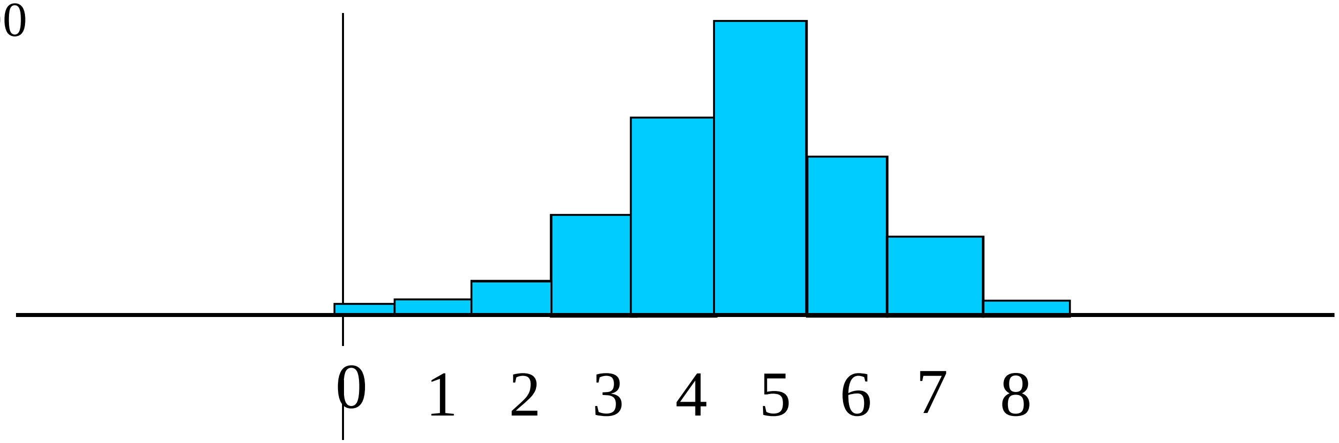
This is asking for a CUMULATIVE PROBABILITY: the probability of 0 getting the disease or 1 or 2 or 3 or 4 or up to 10.

$P(X \leq 10) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) + \ldots + P(X=10) =$

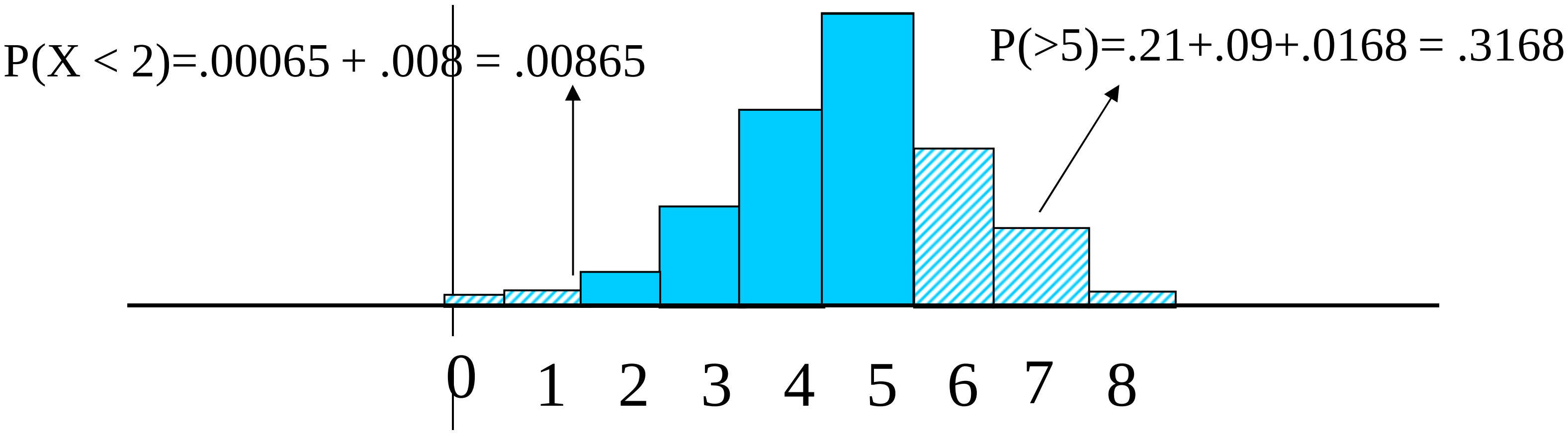
$$\binom{500}{0}(.05)^0(.95)^{500} + \binom{500}{1}(.05)^1(.95)^{499} + \binom{500}{2}(.05)^2(.95)^{498} + \ldots + \binom{500}{10}(.05)^{10}(.95)^{490} < .01$$

Answer

X	P(X)
0	$1(.4)^8 = .00065$
1	$8(.6)^1\ (.4)^7 = .008$
2	$28(.6)^2\ (.4)^6 = .04$
3	$56(.6)^3\ (.4)^5 = .12$
4	$70(.6)^4\ (.4)^4 = .23$
5	$56(.6)^5\ (.4)^3 = .28$
6	$28(.6)^6\ (.4)^2 = .21$
7	$8(.6)^7\ (.4)^1 = .090$
8	$1(.6)^8 = .0168$



Answer, continued



E(X) = 8 (.6) = 4.8

Var(X) = 8 (.6) (.4) =1.92

StdDev(X) = 1.38

Review Question 6

A coin toss can be thought of as an example of a binomial distribution with n=1 and p= 0.5. What are the expected value and variance of a coin toss?

a. 0.5, 0.25
b. 1.0, 1.0
c. 1.5, 0.5
d. 0.25, 0.5
e. 0.5, 0.5

Answer is 0.5 and 0.25

Review Question 7

If I toss a coin 10 times, what is the expected value and variance of the number of heads?

a. 5, 5
b. 10, 5
c. 2.5, 5
d. 5, 2.5
e. 2.5, 10

Answer is 5, 2.5

Poisson Probability Distribution

Poisson Probability Distribution

- When number of trials is large $(n \to \infty)$ and probability of success (or failure) are close to zero or one $(p \to 0\ or\ 1)$. Then the mean value $\lambda = np$ is a moderate value showing the average number of successes in n trials with probability of success p. This is the Poisson distribution.
- Poisson distribution computes the probability in a given time or space interval.

- For example,
  - Average number of flights per minute at Heathrow airport.
  - Average number of mistakes per page by a writer.
  - Average number of Cheetah in Northern areas per square kilometre.
  - Average number of text messages in an hour on your cell phone.

Poisson Distribution –Mean and Variance

- Poisson distribution has the form:

$$P(X = x) = \frac{\lambda^x . e^{-\lambda}}{x!}, \qquad x = 0,1,2, \dots$$

- Mean = E(x) = λ
- Variance = V(x) = λ, therefore, SD= $\sqrt{\lambda}$
- The mean and variance are equal in Poisson distribution.

Poisson Process (Poisson distribution with time/space)

- When time (or space) is involved in a Poisson distribution, then result is called Poisson Process.

$$P(X = x) = \frac{(\lambda t)^x . e^{-\lambda t}}{x!}, \qquad x = 0,1,2,\ldots$$

- Mean = E(x) = λt
- Variance = V(x) = λt, therefore, SD= $\sqrt{\lambda t}$
- The mean and variance are equal in Poisson process.

Questions (to do)

- **Q1.** Defects occur at random in planks of wood with a constant rate of 0.5 per 100 cm length. Jim buys a plank of length 100 cm.
  - (a) Find the probability that Jim's plank contains at most 3 defects.

P(X<=3) = ? You can use Poisson table to calculate this probability

  - (b) Solomon buys 6 planks of length 100 cm. Find the probability that fewer than 2 of Solomon's planks contain at most 3 defects.

Note: If there are 0.5 defects in 100cm then there will be 6x0.5 = 3.0 defects in 600 cm of wood planks.

New mean = 3.0 and then solve the question

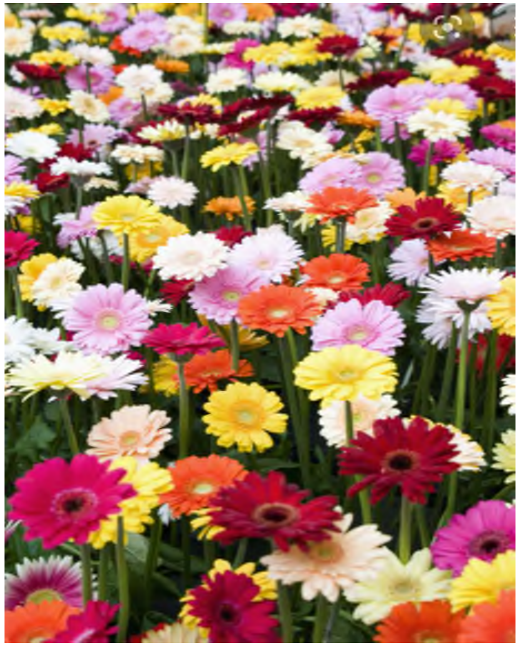
Question - Test Your Understanding

- A shop sells radios at a rate of 2.5 per week.
  a) Find the probability that in a two-week period the shop sells at least 7 radios.
  b) Deliveries of these radios come every 4 weeks. Find the probability of selling fewer than 12 radios in a four-week period.
  c) The manager wishes to make sure that the probability of the shop running out of radios during a four-week period is less than 0.01. Find the smallest number of radios the manager should have in stock immediately after the delivery.

Test your understanding: Poisson Distribution

- A botanist counts the number of daisies, x, in each of 80 randomly selected squares within a field. The results are summarised below.

$$\Sigma x = 295, \qquad \Sigma x^2 = 1386$$



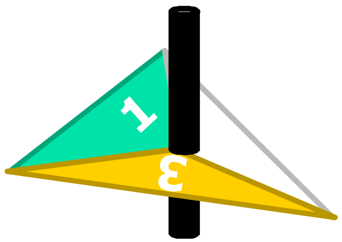
(a) Calculate the mean and the variance of the number of daisies per square for the 80 squares. Give your answers to 2 decimal places.

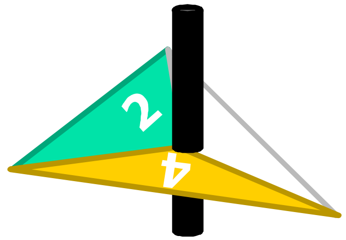
(b) Explain how the answers from part a support the choice of a Poisson distribution as a model.

(c) Using a suitable value for λ, estimate the probability that exactly 3 daises will be found in a randomly selected square.

Adding Random Variables

- Suppose that we had two fair three-sided spinners, each which could be represented using random variables X and Y respectively:
- Then $Z = X + Y$ would represent the distribution of adding each possible outcome from X with each possible outcome from Y:





x	1	2	3
$p(x)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

y	2	3	4
$p(y)$	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$

$z = x + y$	**2**	**3**	**4**
1	3	4	5
2	4	5	6
3	5	6	7

z	3	4	5	6	7
$p(z)$	$\frac{1}{9}$	$\frac{2}{9}$	$\frac{3}{9}$	$\frac{2}{9}$	$\frac{3}{9}$

Each combined outcome has a probability of $\frac{1}{3} \times \frac{1}{3} = \frac{1}{9}$

Poisson Process

✐ If $X \sim Po(\lambda_1)$ and $Y \sim Po(\lambda_2)$ then $X + Y \sim Po(\lambda_1 + \lambda_2)$.
For $X + Y$ to be meaningful, X and Y must represent the same time interval.

- If 5 cars pass per hour in road A and 8 cars pass per hour in road B, how many cars pass per hour in roads A and B combined?

(In class Working)

- **Solve**: If $X \sim Po(3.6)$ and $Y \sim Po(4.4)$ find:

(a) $P(X + Y = 7)$

a

(b) $P(X + Y \leq 5)$

b

Exercise:

- The number of cars passing an observation point in a 5-minute interval is modelled by a Poisson distribution with mean 2. The number of other vehicles passing the observation point in a 15-minute interval is modelled by a Poisson distribution with mean 3. Find the probability that:
  (a) exactly 5 vehicles, of any type, pass the observation point in a 10-minute interval
  (b) more than 8 vehicles, of any type, pass the observation point in a 15-minute interval

- A shop sells radios at a rate of 2.5 per week.
  a) Find the probability that in a two-week period the shop sells at least 7 radios.
  b) Deliveries of these radios come every 4 weeks. Find the probability of selling fewer than 12 radios in a four-week period.

Binomial Approximation using Poisson Distribution

Approximating a Binomial using a Poisson

✐ If $X \sim B(n,p)$ and:

- n is large
- p is small

Then X can be approximated by $Po(np)$

Generally if $np \leq 10$ then a Poisson is suitable enough approximation, but in an exam, use the original Binomial unless instructed to approximate.

$$X \sim B(n,p) \qquad\qquad X \sim Po(np)$$

$$P(X=x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \Longrightarrow \quad P(X=x) = \frac{e^{-\lambda}\lambda^x}{x!}$$

For the Binomial Distribution, the probability involves $\binom{n}{x} = \frac{n!}{x!(n-x)!}$

When n is really large, **calculating $n!$ is really horrid**.

However, the probability function for the Poisson Distribution doesn't involve $n!$ so avoids the problem. Note also that $\lambda = np$ is not too large (as p is small) and so $e^{-\lambda}$ is not too difficult to compute.

Quick test (Binomial approximation using Poisson distribution)

$X \sim B(100, 0.1)$ → ?

$X \sim B(50, 0.5)$ → ?

$X \sim B(40, 0.02)$ → ?

$X \sim B(300, 0.2)$ → ?

Practice Problem: Binomial approx. using Poisson distribution

- 1. You are performing a cohort study. If the probability of developing disease in the exposed group is 0.05 for the study duration, then if you (randomly) sample 500 exposed people,
  - How many do you expect to develop the disease? Give a margin of error (+/- 1 standard deviation) for your estimate.

$X \sim$ binomial (500, .05) with Mean = $E(X) = 500\,(.05) = 25$ and

$Var(X) = 500\,(.05)\,(.95) = 23.75$. so that $SD(X) = 4.87$

Therefore, 25 ± 4.87 is the range of values.

  - What's the probability that **at most** 10 exposed people develop the disease? $P(X \leq 10) = ?$ **(Use the Poisson distribution formula)**

Important questions for practice.
(Binomial probabilities using Poisson distribution)

A call centre also sells the magazine. The probability that a telephone call made by the call centre sells a magazine is 0.05. The call centre telephones 100 people every hour.

(*d*) Using a suitable approximation, find the probability that more than 10 people telephoned by the call centre buy a magazine in a randomly chosen hour.
(3)

The probability of any one letter being delivered to the wrong house is 0.01.
On a randomly selected day Peter delivers 1000 letters.

(*b*) Using a Poisson approximation, find the probability that Peter delivers at least 4 letters to the wrong house.

Give your answer to 4 decimal places.
(3)

A bag contains a large number of counters of which 15% are coloured red. A random sample of 30 counters is selected and the number of red counters is recorded.

(*a*) Find the probability of no more than 6 red counters in this sample.
(2)

A second random sample of 30 counters is selected and the number of red counters is recorded.

(*b*) Using a Poisson approximation, estimate the probability that the total number of red counters in the combined sample of size 60 is less than 13.
(3)

References:

- Following list of references were used for this lecture preparations.
  - Kingston college UK.
  - **Prof. Sher Muhammad Chaudhry and Dr Shahid Kamal, Introduction to Statistical Theory. Ilmi Kitab Khana, Lahore (Part-I and II).**
  - Dr Tahseen jilani old lectures.

End of lesson-02

Thank you