MA203

Special RV

Special RV:

1. Bernoulli RV: X=0,1 and P(X=1)=p, P(X=0)=1-p; 0<p<1

   E(X)=p, Var (X)=p(1-p)

2. Binomial RV: n no. of independent Bernoulli trials and X is the total number of successes in n trials

$P(X=k)={}^{n}C_{k}\, p^{k} (1-p)^{n-k}$ k=0,1,2,….., n; 0<=p<=1

X~B(n,p)

E(X)=np, Var (X)=npq,

Mode: unimodal ([(n+1)p]) or bimodal ((n+1)p, (n+1)p-1)

Example: throwing a die and obtaining even number. X~B(1,1/2)

Theorem 1: Let, Xi (i=1, 2, ….,k) be independent binomial RVs with Xi~B(ni, p)

S=X1+X2+….Xk

S~B(ni+n2+…nk, p)

Example: X1~B(1,1/2), X2~B(1,1/2) both are tossing a coin and obtaining head

Consider, X=tossing a coin twice and obtaining head

X~B(2,1/2)

P(X=0)=1/4, P(X=1)=2/4, P(X=2)=1/4

Example 1: A die is thrown 600 times. Find the lower bound for the prob of getting 80 to 120 sixes.

$X \sim B(600,1/6)$
$P\{|X-\mu|<k\sigma\} >= 1-1/k^2$
$\mu=600*1/6=100$
$Var(X)=600*1/6*5/6=500/6$

$K=2.19$
Lower bound of prob = 19/24

2. Use Chebyshev inequality to determine how many times a fair coin must be tossed in order that the prob will be at least 0.90 that the ratio of the observed no of heads to the no of tosses will lie between 0.4 and 0.6.
Consider, RV X/n, X: no of heads, n: total tosses
$X \sim B(n,1/2)$
$P\{|X/n-E(X/n)|<\varepsilon\} >= 1-(Var(X/n))/\varepsilon^2$
$E(X/n)=1/n.E(X)=1/n(n.1/2)=1/2$, $Var(X/n)=(1/n^2)Var(X)=1/(4n)$
$P\{|X/n-1/2|<k/(4n)\} >= 1-1/k^2$
$P\{|X/n-1/2|<k/(4n)\} >= 0.9$
So, $k=\sqrt{10}$
$N=250$
The coin must be tossed 250 times.

3. Poisson RV: X takes values 0, 1, 2,….. Inf with

P(X=k)=$e^{-\lambda}\lambda^k/k!$, λ>0 DF is $F(x)=\sum_{k=0}^{x}e^{-\lambda}\frac{\lambda^k}{k!}$

X~P(λ)

(To check pmf, $\sum_{k=0}^{\infty}P(X=k)=1$)

Example: number of printing mistakes per page, number of goals in a football match etc..

$$E(X)=\sum_{x=0}^{\infty}xe^{-\lambda}\frac{\lambda^x}{x!}=e^{-\lambda}\sum_{x=1}^{\infty}\frac{\lambda^x}{(x-1)!}=\lambda e^{-\lambda}\sum_{x=1}^{\infty}\frac{\lambda^{x-1}}{(x-1)!}=\lambda$$

Var(X)=E(X2)-(E(X))2 =λ^2+ λ- λ^2= λ

Mode: let, pk=P(X=k)

$$\frac{p_{k-1}}{p_k}=\frac{k}{\lambda}$$

If k< λ, then pk-1<pk

If k> λ, then pk-1>pk

If k= λ, then pk-1=pk

If λ is not an integer, then mode= [λ]

If λ is an integer, then mode = λ, λ-1

If λ<1, mode=0

MGF: $M(s)=e^{\lambda(\exp(s)-1)}$

Theorem 2: Let, $X1, X2, \ldots, Xn$ be independent Poisson RVS with $Xk \sim P(\lambda k)$, $k=1, 2, \ldots, n$.
Then $S \sim P(\lambda 1+ \lambda 2+\ldots \lambda k)$ where $S=X1+X2+\ldots+Xn$
(converse is true)

Theorem 3: Let X, Y be independent Poisson RVs with parameters $\lambda 1$, $\lambda 2$ respectively. Then
$(X|X+Y=n) \sim B$
$P\{X=m|X+Y=n\}=P\{X=m, Y=n-m\}/P\{X+Y=n\}$

$(X+Y) \sim P(\lambda 1+ \lambda 2)$
$(X|X+Y) \sim B(n, \lambda 1 /(\lambda 1+ \lambda 2))$

Theorem 4: if $X \sim P(\lambda)$ and $(Y|X=x) \sim B(x,p)$ then $Y \sim P(\lambda p)$

3. A firm receives on an average 2.5 telephone calls per day during 10 – 10:05 am. Find the probability that on a certain day, the firm receives no call; exactly 4 calls during the same time.
$X \sim P(2.5)$
$P(X=0)=0.0821$
$P(X=4)=0.1336$

4. Uniform RV (discrete) : RV X is said to be have uniform distribution on n points {x1, x2, …. Xn}, if its pmf is P{X=xi}=1/n, i=1,2,…,n
F(x)=x/n
E(X)=1/n(x1+x2+…+xn)=mean

$$Var(X) = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - (E(X))^2$$

If xi=i
E(X)= (n+1)/2
Var (X)= (n^2-1)/12

5. Geometric RV: Suppose, we are interested in the first success, i.e., how many trials are required to realize first success. Then no of trials needed is not fixed and its random numer.
X: no of trials needed to the first success
A: success event
P(X=k)=P(A^c A^c…… A^c A)={(1-p)^(k-1)}.p k=1,2,3,…. Inf
X is said to be geometric RV with P(X=k)=p.q^(k-1), k=1,2,3…… inf

$$P(X > m) = \sum_{k=m+1}^{\infty} q^{k-1}p = pq^m \frac{1}{1-q} = q^m$$

For integers m , n >1

$$P(X > m + n|X > m) = \frac{P(X > m + n)}{P(X > m)} = q^n$$

Given first m trial has no success, conditional prob that first success will appear after an additional n trials depend only on n and not on m – memoryless property

$$E(X) = \sum_{k=1}^{\infty} kq^{k-1}p = \frac{1}{p}$$

$$Var(X) = \sum_{k=1}^{\infty} k^2q^{k-1}p - {}^{1}/_{p^2} = \frac{1-p}{p^2}$$

DF is F(x)=P(X<=x)=$\sum_{u=1}^{x} q^u p = 1 - (1-p)^{x+1}$

MGF is $\mathrm{M}(s) = \sum_{k=1}^{\infty} e^{sk}(1-p)^{k-1}p = \frac{pe^s}{1-(1-p)e^s}$

Theorem-5. if X has a geometric distribution, then for any two non-negative integers m and n, P(X>m+n|X>m)=P(X>n)

- Memoryless distribution

Extend it to the number of trials needed for r successes.
Y: no of trials needed for r successes (Bernoulli trials)

P(Y=k)=P(r-1 successes in k-1 trials and success in k-th trial)
$P(Y=k)=\binom{k-1}{r-1}p^{r-1}(1-p)^{k-1-r+1}p=\binom{k-1}{r-1}p^{r}(1-p)^{k-r},$ k=r, r+1, r+2, ……. Inf
RV Y is called negative binomial RV.

6. Uniform RV (continuous)- an RV X is said to be uniformly distributed in the interval (a,b), if
f(x)=1/(b-a), a<=x<=b
=0, otherwise
X~U(a,b)
F(x)=0, x<a
=(x-a)/(b-a), a<=x<b
=1, x>=b

E(X)=
Var (X)=

MGF is M(s)=

E(X)= (a+b)/2
Var (X)= (b-a)^2/12

MGF is M(s)= (exp(sb)-exp(sa))/(s(b-a))

7. Gaussian RV:
X can take values –inf to +inf
An RV X is said Gaussian or Normal RV with parameters μ and σ^2 if

$$f_X(x) = \frac{1}{\sqrt{(2\pi)}\sigma} e^{-(x-\mu)^2 / 2\sigma^2} \;, -\infty \le x \le \infty$$

X~N(μ , σ^2)
Bell-shaped curve, symmetric around parameter μ

DF is $F_X(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{(2\pi)}\sigma} e^{-(y-\mu)^2 / 2\sigma^2} dy$

F(x) is available in tabular form
E(X)=μ, Var(X)= σ^2

Let Y is another RV as Y=(X- μ)/σ
E(Y)=E[(X- μ)/σ]=(1/σ)[E(X)- μ]=0, Var(Y)=(1/σ^2)Var(X)=1
Y~N(0,1) – standard normal RV

MGF is $M(s) = E(e^{sX}) = e^{(\mu s+0.5(\sigma s)^{\wedge}2)}$, for all real values of s

Moments of all order exist and can be computed from MGF.

$\therefore Var(X) = E(X$

<u>Theorem-8.</u> Let, $X_1, X_2, \ldots, X_n$ be independent RVs with $X_k \sim N(\mu_k, \sigma_k^2)$, $k = 1, 2, \ldots, n$; then

$$S_n \sim N\left(\sum_{k=1}^{n} \mu_k, \sum_{k=1}^{n} \sigma_k^2\right) \text{ where } S_n = \sum_{k=1}^{n} X_k$$

Corollary — if X_i's are iid $N(\mu, \sigma^2)$ RVs, then

$$S_n \sim N(n\mu, n\sigma^2) \text{ where } S_n = \sum_{k=1}^{n} X_k$$

also $$\frac{1}{n} S_n \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

if X_i's are iid $N(0,1)$ RVs, then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i = \frac{1}{\sqrt{n}} S_n \sim N(0, 1)$$

Theorem 9. Let, X and Y be independent RVs. Then (X+Y) follows Normal distribution iff X and Y are both normal.

Theorem-11. if $X \sim N(0,1)$ then

$$X^2 \sim \chi^2(1) \quad \text{Chi-square distribution}$$

if $X \sim N(\mu, \sigma^2)$ then $Z = \left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi^2(1)$

Asymptotic approximation —

Let, $X \sim B(n, p)$

$$\therefore P(k_1 \leq X \leq k_2) = \sum_{k=k_1}^{k_2} p_n(k) = \sum_{k=k_1}^{k_2} \binom{n}{k} p^k q^{n-k} \quad \text{——} ①$$

$$= \sum_{k=k_1}^{k_2} \frac{n!}{k!(n-k)!} p^k q^{n-k}$$

since $\binom{n}{k}$ grows rapidly with n, it is difficult to compute for large n. In this context, two approximations are used. (De Moivre-Laplace theorem)

since $\binom{n}{k}$ grows rapidly ... for large n. In this context, two approxima... (De Moivre-Laplace theorem)

1) Normal approximation (De Moivre-Laplace theorem)

Suppose, $n \to \infty$ with p held fixed. Then for k in the $\sqrt{npq}$ neighbourhood of np, we can approximate,

$$\binom{n}{k} p^k q^{n-k} \simeq \frac{1}{\sqrt{2\pi npq}} e^{-(k-np)^2/2npq} \quad , p+q=1$$

ratio of 2 sides tends to 1 as $n \to \infty$. So, if k_1 and k_2 are within or around the neighbourhood of the interval $(np - \sqrt{npq}, np + \sqrt{npq})$, then the summation in ① can be approximated as

$$P(k_1 \leq X \leq k_2) = \int_{k_1}^{k_2} \frac{1}{\sqrt{2\pi npq}} e^{-\frac{(k-np)^2}{2npq}} dk$$

$$\therefore X \sim B(n,p) \simeq X \sim N(np, npq) \qquad p+q=1.$$

2) <u>Poisson approximation</u> —

Normal approximation is valid only if p is fixed, that is only if $np \gg 1$, $npq \gg 1$ and p is fixed. This approximation deteriorates as $p \to 0$ or $p \to 1$ and it completely fails if $p = 0, q = 1$ or $p = 1, q = 0$

So, if $n \to \infty$, $p \to 0$ but np is finite, a fixed number, then Poisson approximation is used.

$$\binom{n}{k} p^{k} q^{n-k} \xrightarrow[n \to \infty]{} e^{-\lambda} \frac{\lambda^{k}}{k!} \quad , k = 0, 1, 2, \ldots$$

$n \to \infty$, $p \to 0$, s.t. $np \to \lambda$

$$\therefore X \sim B(n, p) \longrightarrow X \sim P(np)$$

$$\therefore P(k_1 \leq X \leq k_2) = e^{-np} \sum_{k=k_1}^{k_2} \frac{(np)^{k}}{k!}$$

8. Exponential RV: if occurrences of events over non-overlapping intervals are independent, like arrival times of telephone calls, bus arrival time at a bus stop; then waiting time distribution of these events can be shown to be exponential.
X is exponential RV with parameter λ if its pdf is

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x}, x \geq 0 \\ 0, otherwise \end{cases}$$

DF is $F_X(t) = P(X \leq t) = \int_0^t \lambda e^{-\lambda x} dx = 1 - e^{-\lambda t}$

$P(X>t)=e^{-\lambda t}$

$E(X)=1/\lambda$, $Var(X)=1/\lambda^2$

$$E(X) = \int_0^\infty x\lambda e^{-\lambda x} dx \qquad Var(X) = \int_0^\infty x^2 \lambda e^{-\lambda x} dx - {}^1/_{\lambda^2}$$

MGF is $M(S)=E(e^{sx})= \lambda/(\lambda-s)$, $s< \lambda$

Memoryless property:
$P(X>t+s|X>s)=P(X>t)$, $s,t>=0$

Other distributions: Rayleigh, Nakagami – used in communication
Chi-square, Beta, Cauchy, Weibull, Laplace etc.

Numerical problems:

1. When sending messages over a network, there is a chance that bits will be corrupted. Hamming code allows for a 4-bit code to be encoded as 7 bits with advantage that if at most one bit is corrupted, it can be perfectly reconstructed. If prob of any bit being lost is 0.1, then how does the reliability will change when we use Hamming code?

2. The mean of a Binomial RV is 4 and sd is 3. T/F?

3. Let, X and Y be independent RVs with pmf $P(X=k)=p_k$ and $P(Y=k)=q_k$, $k=0,1,2,\ldots\ldots$

If $P(X=k|X+Y=t)={}^tC_k a^k(1-a)^{t-k}$, find p_k, q_k.

4. A fair die is rolled repeatedly. Compute the probability of event A that a 2 will show up before a 5.

5. Suppose the life length of an appliance has an exponential distribution with mean 10 years. A used appliance is bought by someone. What is the prob that it will not fail in the next 5 years?

6. The local authorities in a city install 10000 bulbs in the streets. If these lamps have an average life of 1000 burning hours with an sd of 200 hours, assuming normality, what number of bulbs might be expected to be fail i) in the first 800 burning hours, ii) between 800 and 1200 burning hours? After what period of burning hours would you expect that a) 10% of the bulbs will fail; b) 10% of the bulbs would be still burning?

Standard Normal Probabilities

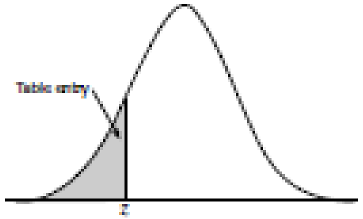
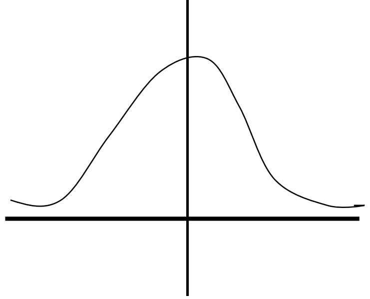


Table entry for z is the area under the standard normal curve to the left of z.

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
−3.4	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0002
−3.3	.0005	.0005	.0005	.0004	.0004	.0004	.0004	.0004	.0004	.0003
−3.2	.0007	.0007	.0006	.0006	.0006	.0006	.0006	.0005	.0005	.0005
−3.1	.0010	.0009	.0009	.0009	.0008	.0008	.0008	.0008	.0007	.0007
−3.0	.0013	.0013	.0013	.0012	.0012	.0011	.0011	.0011	.0010	.0010
−2.9	.0019	.0018	.0018	.0017	.0016	.0016	.0015	.0015	.0014	.0014
−2.8	.0026	.0025	.0024	.0023	.0023	.0022	.0021	.0021	.0020	.0019
−2.7	.0035	.0034	.0033	.0032	.0031	.0030	.0029	.0028	.0027	.0026
−2.6	.0047	.0045	.0044	.0043	.0041	.0040	.0039	.0038	.0037	.0036
−2.5	.0062	.0060	.0059	.0057	.0055	.0054	.0052	.0051	.0049	.0048
−2.4	.0082	.0080	.0078	.0075	.0073	.0071	.0069	.0068	.0066	.0064
−2.3	.0107	.0104	.0102	.0099	.0096	.0094	.0091	.0089	.0087	.0084
−2.2	.0139	.0136	.0132	.0129	.0125	.0122	.0119	.0116	.0113	.0110
−2.1	.0179	.0174	.0170	.0166	.0162	.0158	.0154	.0150	.0146	.0143
−2.0	.0228	.0222	.0217	.0212	.0207	.0202	.0197	.0192	.0188	.0183
−1.9	.0287	.0281	.0274	.0268	.0262	.0256	.0250	.0244	.0239	.0233
−1.8	.0359	.0351	.0344	.0336	.0329	.0322	.0314	.0307	.0301	.0294
−1.7	.0446	.0436	.0427	.0418	.0409	.0401	.0392	.0384	.0375	.0367
−1.6	.0548	.0537	.0526	.0516	.0505	.0495	.0485	.0475	.0465	.0455
−1.5	.0668	.0655	.0643	.0630	.0618	.0606	.0594	.0582	.0571	.0559
−1.4	.0808	.0793	.0778	.0764	.0749	.0735	.0721	.0708	.0694	.0681
−1.3	.0968	.0951	.0934	.0918	.0901	.0885	.0869	.0853	.0838	.0823
−1.2	.1151	.1131	.1112	.1093	.1075	.1056	.1038	.1020	.1003	.0985
−1.1	.1357	.1335	.1314	.1292	.1271	.1251	.1230	.1210	.1190	.1170
−1.0	.1587	.1562	.1539	.1515	.1492	.1469	.1446	.1423	.1401	.1379
−0.9	.1841	.1814	.1788	.1762	.1736	.1711	.1685	.1660	.1635	.1611
−0.8	.2119	.2090	.2061	.2033	.2005	.1977	.1949	.1922	.1894	.1867
−0.7	.2420	.2389	.2358	.2327	.2296	.2266	.2236	.2206	.2177	.2148
−0.6	.2743	.2709	.2676	.2643	.2611	.2578	.2546	.2514	.2483	.2451
−0.5	.3085	.3050	.3015	.2981	.2946	.2912	.2877	.2843	.2810	.2776
−0.4	.3446	.3409	.3372	.3336	.3300	.3264	.3228	.3192	.3156	.3121
−0.3	.3821	.3783	.3745	.3707	.3669	.3632	.3594	.3557	.3520	.3483
−0.2	.4207	.4168	.4129	.4090	.4052	.4013	.3974	.3936	.3897	.3859
−0.1	.4602	.4562	.4522	.4483	.4443	.4404	.4364	.4325	.4286	.4247
−0.0	.5000	.4960	.4920	.4880	.4840	.4801	.4761	.4721	.4681	.4641



APPENDIX

TABLE I

Area Under Standard Normal Curve

(Proportion of area under standard normal curve between the ordinates at $z = 0$ and given values of z)

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990
3.1	.4990	.4991	.4991	.4991	.4992	.4992	.4992	.4992	.4993	.4993
3.2	.4993	.4993	.4994	.4994	.4994	.4994	.4994	.4995	.4995	.4995
3.3	.4995	.4995	.4995	.4996	.4996	.4996	.4996	.4996	.4996	.4997
3.4	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4997	.4998

7. Six coins are tossed 6400 times. Using Poisson approximation, find the approximate prob of getting six heads r times.

8. A fair die is thrown 600 times. Find the prob of getting 80 to 120 sixes.

9. In a normal distribution, 8% of the items are under 50 and 10% are over 60. Find the mean and sd. [given, $\int_x^{\infty} \frac{1}{\sqrt{2\pi}} e^{-z^2/2} dz = 0.08 \ or \ 0.1 \ as \ x = 1.4 \ or \ 1.28 \ respectively.$]