Unit - 3 Probability

Random Variable

→ $P(X) \Rightarrow \quad X = \{1, 2, 3, 4, 5, 6\}$

$\downarrow$ each: $\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}$

$P_i = P(X = x_i) \Rightarrow$ Probability Distribution (or)
Probability Mass Function

Cummulative Distribution Function

→ $P(X \leq x_i) = F(x_i)$

ex: $P(X \leq 3) = P(X = 1) + P(X = 2) + P(X = 3)$

Mean, Variance, Standard Deviation

→ $X = \{x_1, x_2, \dots x_n\}$ &

Prob. mass func^n are P_i, $i = 1, 2, 3, \dots n$

Then,

$$\text{mean } (\mu) = \sum_{i=1}^{n} x_i P_i$$

$$\text{variance } (\sigma^2) = \sum_{i=1}^{n} x_i^2 P_i - \mu^2$$

$$\text{Standard deviation} = \sqrt{\sigma^2} = \sigma$$

Continuous Distribution Frequency

→ If x takes every value in an interval, its called continuous distribution

→ Prob. of x lies in the interval $x - \frac{dx}{2}$ & $x + \frac{dx}{2}$

→ $P\left(x - \frac{dx}{2} < x < x + \frac{dx}{2}\right) = f(x)\,dx$

→ If $f(x) = \phi(x)$ be a density variable function denoted for variable x, in interval $[a, b]$

$$f(x) = \begin{cases} 0, & x < a \\ \phi(x), & a \leq x \leq b \\ 0, & x > b \end{cases} \Rightarrow \int_{-\infty}^{\infty} f(x)\,dx = 1$$

$$\Rightarrow \int_{a}^{b} f(x)\,dx = 1$$

$\rightarrow \quad F(x) = P(x \leq x_i)$

$$= \int_{-\infty}^{\infty} f(x)\,dx$$

$\rightarrow \quad F'(x) = f(x)$

$\rightarrow \quad F(-\infty) = 0$

$\rightarrow \quad F(\infty) = 1$

$\rightarrow \quad P(a \leq x \leq b) = \int_a^b f(x)\,dx = \int_{-\infty}^{b} f(x)\,dx - \int_{-\infty}^{a} f(x)\,dx$

$$= F(b) - F(a)$$

Mean & Variance for continuous distribution functions

$\rightarrow \quad \mu = \int_{-\infty}^{\infty} x . f(x)\,dx \qquad$ (Mean)

$\rightarrow \quad \sigma^2 = \int_{-\infty}^{\infty} x^2 f(x) - \mu^2 \quad$ (variance)

Discrete
- → Bernoulli's
- → Binomial
- → Poisson

Continuous
- → Normal
- → Exponential

Bernoulli's Distribution

→ Single Trial
→ Only 2 Outcomes — Success & Failure

$$\text{Success} \downarrow x = 1 \qquad \text{Failure} \downarrow x = 0$$

$P(x = 1) = P$

$P(x = 0) = 1 - P$

$\mu = \Sigma x_i P_i = P$

$\sigma^2 = \Sigma x_i^2 P_i - \mu^2 = P(1-P)$

Binomial Distribution

→ If n bernoulli's trials are conducted $(n > 1)$,

i) trials are independant

ii) Each trial has same success prob.

and, x is no. of successful trials,

n trials

p is prob. of x successful

Then, $E(x)$ (Mean) $= np$

$V(x)$ (Variance) $= np(1-p) = npq$

→ If a series of independant bernoulli's trials are performed, such that $p \to$ prob. of success &

$q \to$ prob. of failure,

$$P(x = r) = {}^{n}C_r \cdot p^r \cdot q^{n-r}$$

$\mu = np$

$\sigma^2 = npq$

Poisson's Distribution

→ n tends to ∞
→ p tends to 0

Then, $np = \lambda \Rightarrow p = \frac{\lambda}{n}$ & $q = 1 - \frac{\lambda}{n}$

$$P(X = x) = {}^{n}C_{x}\, p^{x} q^{n-x}$$

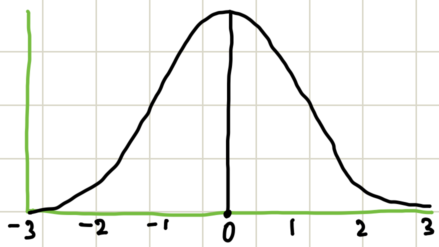
$$= \frac{\lambda^{x} e^{-\lambda}}{x!}$$

$$\mu = \lambda$$

$$\sigma^{2} = \lambda$$

Normal Distribution

→ It is a continuous prob. distribution of continuous random variable
→ $f(x) = P(X = x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^{2}}$



→ This gives a bell curve
→ Mean, median & mode coincide
→ Area under curve $= 1$

$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

$$P(X_1 \leq x \leq X_2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{X_1}^{X_2} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^{2}} dx$$

This formula not really necessary

Note :

68% area lies b/w $\mu - \sigma$ to $\mu + \sigma$
95% area lies b/w $\mu - 2\sigma$ to $\mu + 2\sigma$
99.73% area lies b/w $\mu - 3\sigma$ to $\mu + 3\sigma$

Z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

Standard Normal Distribution

→ Normal distribution for each $\mu = 0$ & $\sigma = 1$

$$z = \frac{x - \mu}{\sigma}$$

z : standard normal variate

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(z)^2}$$

$$E(x) = \mu = \int_{-\infty}^{\infty} \frac{x}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx$$

$$V(x) = \sigma^2 = \int_{-\infty}^{\infty} \frac{x^2}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} dx$$

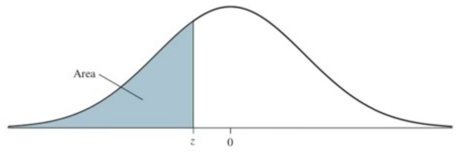
Evaluation of $P(a \leq x \leq b)$

$x = a \Rightarrow z_1 = \frac{a - \mu}{\sigma}$

$x = b \Rightarrow z_2 = \frac{b - \mu}{\sigma}$

$$P(z_1 \leq z \leq z_2) = P(\infty \leq z \leq z_2) - P(-\infty \leq z \leq z_1)$$
$$= P(z \leq z_2) - P(z \leq z_1)$$

Table A.2 Cumulative Normal Distribution



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−3.7 or less	.0001									
−3.6	.0002	.0002	.0001	.0001	.0001	.0001	.0001	.0001	.0001	.0001
−3.5	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002	.0002
−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

Properties:

→ $P(-\infty \leq z \leq \infty) = 1$

→ $P(z \leq 0) = 0.5$ & $P(z \geq 0) = 0.5$

→ If $c > 0$

$P(-c \leq z \leq 0) = P(0 \leq z \leq c)$

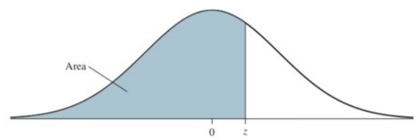
→ $P(-c \leq z \leq c) = 2P(z \leq c)$

→ $P(|z| \geq c) = 1 - P(|z| \leq c)$

$= 1 - P(-c \leq z \leq c)$

$= 1 - 2P(z \leq c)$

Table A.2 Cumulative Normal Distribution (continued)



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962	.8980	.8997	.9015
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131	.9147	.9162	.9177
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279	.9292	.9306	.9319
1.5	.9332	.9345	.9357	.9370	.9382	.9394	.9406	.9418	.9429	.9441
1.6	.9452	.9463	.9474	.9484	.9495	.9505	.9515	.9525	.9535	.9545
1.7	.9554	.9564	.9573	.9582	.9591	.9599	.9608	.9616	.9625	.9633
1.8	.9641	.9649	.9656	.9664	.9671	.9678	.9686	.9693	.9699	.9706
1.9	.9713	.9719	.9726	.9732	.9738	.9744	.9750	.9756	.9761	.9767
2.0	.9772	.9778	.9783	.9788	.9793	.9798	.9803	.9808	.9812	.9817
2.1	.9821	.9826	.9830	.9834	.9838	.9842	.9846	.9850	.9854	.9857
2.2	.9861	.9864	.9868	.9871	.9875	.9878	.9881	.9884	.9887	.9890
2.3	.9893	.9896	.9898	.9901	.9904	.9906	.9909	.9911	.9913	.9916
2.4	.9918	.9920	.9922	.9925	.9927	.9929	.9931	.9932	.9934	.9936
2.5	.9938	.9940	.9941	.9943	.9945	.9946	.9948	.9949	.9951	.9952
2.6	.9953	.9955	.9956	.9957	.9959	.9960	.9961	.9962	.9963	.9964
2.7	.9965	.9966	.9967	.9968	.9969	.9970	.9971	.9972	.9973	.9974
2.8	.9974	.9975	.9976	.9977	.9977	.9978	.9979	.9979	.9980	.9981
2.9	.9981	.9982	.9982	.9983	.9984	.9984	.9985	.9985	.9986	.9986
3.0	.9987	.9987	.9987	.9988	.9988	.9989	.9989	.9989	.9990	.9990
3.1	.9990	.9991	.9991	.9991	.9992	.9992	.9992	.9992	.9993	.9993
3.2	.9993	.9993	.9994	.9994	.9994	.9994	.9994	.9995	.9995	.9995
3.3	.9995	.9995	.9995	.9996	.9996	.9996	.9996	.9996	.9996	.9997
3.4	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9997	.9998
3.5	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998	.9998
3.6	.9998	.9998	.9999	.9999	.9999	.9999	.9999	.9999	.9999	.9999
3.7 or more	.9999									

Exponential Distribution

→ Continuous Distribution

→ Inverse of Poisson Distribution

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

$$F(x) = \begin{cases} 1 - e^{-\lambda x}, & x > 0 \\ 0, & x \leq 0 \end{cases}$$

$$\mu = \frac{1}{\lambda} \quad \& \quad \sigma^2 = \frac{1}{\lambda^2}$$

Relation b/w Exponential & Poisson Distribution

→ If an event follows poisson process with $\mu = \lambda$ & T represent waiting time from any starting point until next occurs, then T is exponentially distributed

$X \sim \text{Poisson}(\lambda)$

$T \sim \text{Exp}(\lambda)$

Normal Approximation to Binomial Distribution

Central Limit Theorem

→ If $x_1, x_2, x_3, \ldots x_n$ etc., be random sample from a population with mean μ & variance σ^2

Let $S_n = x_1 + x_2 + \ldots + x_n$ be sum of sample objects

& $\bar{x} = \dfrac{x_1 + x_2 + \ldots + x_n}{n}$ be sample mean

then, $S_n \sim N(n\mu, n\sigma^2)$ approximately

$\bar{x} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$

(if n is large)

→ If $X \sim \text{Bin}(n, p)$ then

$X = \underbrace{Y_1 + Y_2 + Y_3 + \ldots + Y_n}$

→ Samples from Bernoulli distribution

→ Sum of sample observations

Then by CLT, $X \sim N(np, np(1-p))$

where $\mu = np$ & $\sigma^2 = np(1-p)$

→ If $X \sim \text{Bin}(n, p)$ if $np \geq 10$ & $np(1-p) \geq 1$

then $X \sim N(np, np(1-p))$ approximately

Note : imp af

- $P(X \geq x) = P\left(Z > \frac{x - 0.5 - \mu}{\sigma}\right)$
- $P(X > x) = P\left(Z > \frac{x + 0.5 - \mu}{\sigma}\right)$
- $P(X \leq x) = P\left(Z < \frac{x + 0.5 - \mu}{\sigma}\right)$
- $P(X < x) = P\left(Z < \frac{x - 0.5 - \mu}{\sigma}\right)$