Random Variables

Example: genders of 3 children

3 BBB	1 GBG	1 GGB
3 GGG	1 BBG	
1 BGB	1 BGG	1 GBB

Y = Difference in no. of each gender

Y(BBB) = 3
Y(GBG) = 1
Y(GGG) = 3
……

$\{X=2\}=\{BBG, BGB, GBB\}$ $\Pr(X=2)=\Pr(\{BBG,BGB,GBB\})=0.375$

$\{Y=2\}=\{\ \}=\phi$ $\Pr(Y=2)=\Pr(\phi)=0$

Probability Mass Function

$p(0)$ = Pr(X = 0) = Pr({GGG}) = 0.125
$p(1)$ = Pr(X = 1) = Pr({BGG, GBG, GGB}) = 0.375
$p(2)$ = Pr(X = 2) = Pr({BBG, BGB, GBB}) = 0.375
$p(3)$ = Pr(X = 3) = PR({BBB}) = 0.125

Space of X : $N=\{0,1,2,3\}$

pmf of X : $p(x)=\binom{3}{x}(0.5)^3 \quad , x=0,1,2,3$

$p(2)=0.375$

$p(3)=0.125$

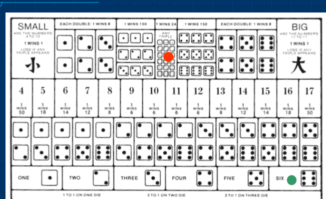
Properties of pmf

1. $p(x)\geq 0$
2. $\sum_{x\in N} p(x)=1$
3. $\Pr(X\in A)=\sum_{x\in A} p(x)$

Mathematical Expectation

Example: Sic Bo ( )

Toss 3 fair dice



E(gain by ●) = -30.56%

E(gain by ●) = -7.87%

Best strategy : don't bet!

Bet option 1 : win $24 if same number on three dice, otherwise lose $1

Bet option 2 : win $k if there are k dice faced up as six, otherwise lose $1

Mathematical Expectation

Expected value of random variable X $\quad E(X)=\sum_{x\in N} xp(x)$

Expected value of X^2 $\quad E(X^2)=\sum_{x\in N} x^2 p(x)$

Expected value of log(X) $\quad E(\log X)=\sum_{x\in N}(\log x)p(x)$

$E(5)=5$ $\quad E(3.7)=3.7$

$E(5X+3X^2-10\sqrt{X})=5E(X)+3E(X^2)-10E(\sqrt{X})$

$E(X^2)\neq E(X)^2$ $\quad E(\log X)\neq \log E(X)$ ……

Mean and Variance

Example: investment

Return on Stock	-10%	0%	10%	20%
Probability	0.1	0.2	0.4	0.3

$E(S)=9\%$

$Var\ (S)=89\ \%^2$

Return on Bonds	6%	8%	10%
Probability	0.2	0.6	0.2

$E(B)=8\%$

$Var\ (B)=1.6\ \%^2$

Population Mean : $\mu=E(X)$

$E[(S\ E[(B-8)^2]=(-2\ldots)^2(0.6)+(2)^2(0.2)=1.6\ldots.3)=89$

Population Variance : $\sigma^2=Var(X)=E[(X-\mu)^2]$
$=E(X^2)-\mu^2$

Mean and Variance

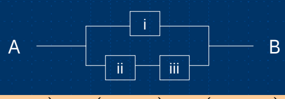
	Sample Use relative frequency $\frac{f}{n}$	Population Use pmf $p(x)$
Mean	$\bar{X}=\sum x\left(\frac{f}{n}\right)$	$\mu=\sum xp(x)$
Variance	$SD^2=\sum(x-\bar{X})^2\left(\frac{f}{n}\right)$	$\sigma^2=\sum(x-\mu)^2 p(x)$

$E(a\ \overline{aX+c}=a\bar{X}+c\ c$

$Var(a\ S^2_{aX+c}=a^2S^2_X\ \ r(X)$

Mean and Variance

Example : water flow



$\Pr(i \text{ open}) = \Pr(ii \text{ open}) = \Pr(iii \text{ open}) = 0.8$

Y = no. of open paths from A to B　　　$Y \in \{ 0,1,2 \}$

$p(2) = 0.512$).8 $p(0) = 0.072$ 1 $- p(1) = 0.416$ 0) $- p(2)$

$E(Y) = 1.44$).072 $E(Y^2) = 2.464$).0 $Var(Y) = 0.3904$ $(1.44)^2$ .512

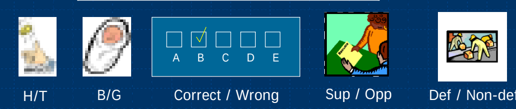
$X = 80Y$

$E(X) = 80E(Y) = 115.2$　　$Var(X) = 80^2 Var(Y) = 2498.56$

Binomial Distribution

Bernoulli Experiment/Trial

Possible outcome	Probability
Success	π
Fail	1 - π



X = No. of successes in n independent Bernoulli trials

X is said to have a b $X \sim b(n, \pi)$ ution with n trials and success probability π

Binomial Distribution

Example: n = 4, π = 0.2

Value of X	0	1	2	3	4
Outcome	FFFF	SFFF FSFF FFSF FFFS	SSFF SFSF SFFS FSSF FSFS FFSS	SSSF SSFS SFSS FSSS	SSSS
Prob.	$(0.8)^4$	$(0.2)(0.8)^3$	$(0.2)^2(0.8)^2$	$(0.2)^3(0.8)$	$(0.2)^4$
No. of combinations	${}_4C_0$	${}_4C_1$	${}_4C_2$	${}_4C_3$	${}_4C_4$

$p(x) = \Pr(X = x) = {}_4C_x (0.2)^x (0.8)^{4-x}$, $x = 0,1,2,3,4$

$p(0) = 0.4096$　$p(1) = 0.4096$　$p(2) = 0.1536$　$p(3) = 0.0256$　$p(4) = 0.0016$

Binomial Distribution

$X \sim b(n, \pi)$

$p(x) = \Pr(X = x) = {}_nC_x \pi^x (1-\pi)^{n-x}$, $x = 0,1,\ldots,n$

$1 = (\pi + (1-\pi))^n = \sum_{x=0}^{n} {}_nC_x \pi^x (1-\pi)^{n-x}$ —— Binomial Theorem

1

$E(X) = n\pi$,　$Var(X) = n\pi(1-\pi)$

$(1-\pi)^n$　${}_nC_1\pi(1-\pi)^{n-1}$　${}_nC_2\pi^2(1-\pi)^{n-2}$　……　π^n

X = 0　X = 1　X = 2　X = n

Binomial Distribution

Example: 50 Multiple choice questions

Correct: 2 marks　　Incorrect: -1 mark

X = no. of correct answers by pure guess

$X \sim b(50, 0.2)$　$E(X) = (50)(0.2) = 10$　$Var(X) = (50)(0.2)(0.8) = 8$

$\Pr(X = 15) = p(15) = {}_{50}C_{15}(0.2)^{15}(0.8)^{35} = 0.02992$

$Y = 2 \times X + (-1) \times (50 - X) = 3X - 50$ .9393 = 0.0607

$E(Y) = E(3X - 50) = 3E(X) - 50 = 3 \times 10 - 50 = -20$

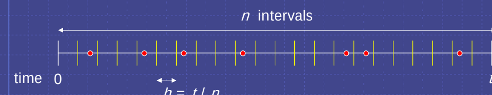
$Var(Y) = Var(3X - 50) = 9Var(X) = 9 \times 8 = 72$

$\Pr(Y < 0) = \Pr(3X - 50 < 0) = \Pr(X \le 16) = 0.9856$

$\Pr(Y \ge 40) = \Pr(3X - 50 \ge 40) = \Pr(X \ge 30) \approx 0$

Poisson Distribution

Example: no. of customers arrived before time t　$N(t)$



Assumptions :

1. Independence of non-overlapping time intervals
2. Customers won't cluster in a short interval
3. Probability of arrival in a short interval ∝ length of interval

$$\Pr(N(t) = x) = \frac{e^{-\lambda t}(\lambda t)^x}{x!} \quad , x = 0,1,2,\ldots$$

Poisson Distribution

$Y \sim \wp(\theta)$

$$p(y)=\Pr(Y=y)=\frac{e^{-\theta}\theta^{y}}{y!}\quad ,\ y=0,1,2,\ldots$$

$$E(Y)=Var(Y)=\theta$$

Poisson process : $N(t)$ = no. of occurrences within [0, t]

λ = average number of occurrences per unit time

$$N(t)\sim\wp(\lambda t)$$

Poisson Distribution

Example : average 3 phone calls per hour

time unit = 1 hour, λ = 3

$$\Pr(\text{2 phone calls in one hour})=\Pr(N(1)=2)=\frac{e^{-3}3^{2}}{2!}=0.224$$

$$\Pr(\text{at least 2 phone calls in one hour})=\Pr(N(1)\geq 2)$$
$$=1-\frac{e^{-3}3^{0}}{0!}-\frac{e^{-3}3^{1}}{1!}=0.801$$

$$\Pr(\text{less than 8 phone calls in 2 hours})=?$$

$$N(2)\sim\wp(6)\qquad \Pr(N(2)\leq 7)=\sum_{y=0}^{7}\frac{e^{-6}6^{y}}{y!}=0.744$$

Poisson Approximation to Binomial

If n is large and π is small,
binomial can be approximated by Poisson.

$$b(n,\pi)\cong\wp(n\pi)$$

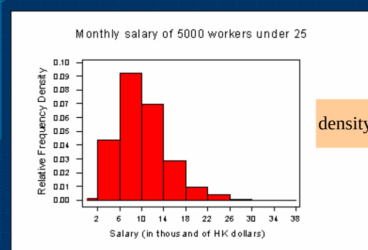
$$p(x)={}_{n}C_{x}\pi^{x}(1-\pi)^{n-x}\approx\frac{e^{-n\pi}(n\pi)^{x}}{x!}$$

Example : $X\sim b(8000,0.001)$

$$\Pr(X\leq 7)\approx\sum_{x=0}^{7}\frac{e^{-8}8^{x}}{x!}=0.4530\text{)}.999)^{8000-x}$$

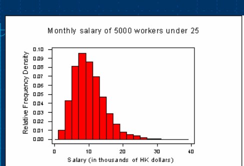
Continuous Distribution

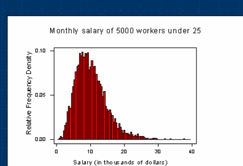
Recall the relative frequency histogram



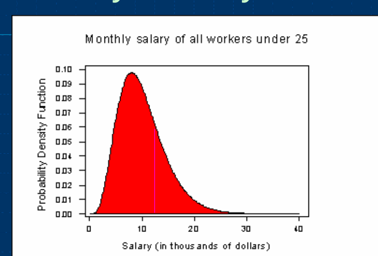
$$\text{density}=\frac{\text{relative frequency}}{\text{class width}}$$

Continuous Distribution



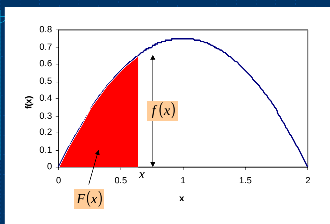


Probability Density Function



$$\text{density}=\text{(ct}\ f(x)=\lim_{t\to 0}\frac{\Pr(X\leq x+t)-\Pr(X\leq x)\text{e r.f. up to }x)}{t}$$

Distribution Function



$f(x)=\frac{3}{4}x(2-x) \quad ,0<x<2$

$\Pr(X \le 0.5)=0.15625$

$\Pr(1\le X \le 1.5)=0.34375$

……

Distribution function $F(x)=\frac{1}{4}(3x^2-x^3)$

$\Pr(X \le 0.5)=F(0.5)$ $\quad$ $\Pr(1\le X \le 1.5)=F(1.5)-F(1)$

Mean and Variance

	Discrete pmf $p(x)$	Continuous pdf $f(x)$
Mean	$\mu=\sum xp(x)$	$\mu=\int xf(x)dx$
Variance	$\sigma^2=\sum(x-\mu)^2 p(x)$	$\sigma^2=\int(x-\mu)^2 f(x)dx$

$Var(X)=E(X^2)-\mu^2=\int x^2 f(x)dx-\mu^2$

Exponential Distribution

Example: $N(t)$ = no. of customers arrived before time t

time 0 $\quad t$

Assumptions :

1. Independence of non-overlapping time intervals
2. Customers won't cluster in a short interval
3. Probability of arrival in a short interval ∝ length of interval

$$\Pr(N(t)=y)=\frac{e^{-\lambda t}(\lambda t)^x}{x!} \quad , y=0,1,2,...$$

Exponential Distribution

Example: T = waiting time until the first customer arrive

$T \quad T$

time 0 $\quad t$

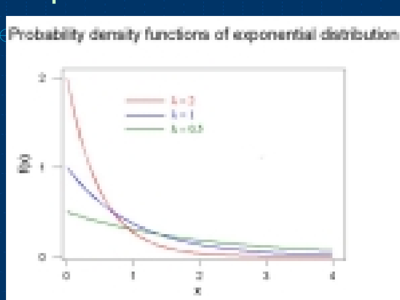
$T>t \Leftrightarrow N(t)=0$ $\quad$ $T\le t \Leftrightarrow N(t)>0$

$F(t)=1-e^{-\lambda t}$ $\quad$ 0)

$f(t)=F'(t)=\lambda e^{-\lambda t}$

$T \sim Exp(\lambda)$

Exponential Distribution



$E(T)=\frac{1}{\lambda}$

$Var(T)=\frac{1}{\lambda^2}$

Exponential Distribution

Example : Average 2 failures of a machine in one day

Assume Poisson process

time unit = day $\quad$ λ = 2

T = time till the first failure $\quad$ $T \sim Exp(2)$

$E(T)=\frac{1}{2}$ $\quad$ $\Pr(T>4)=1-F(4)=1-(1-e^{-2(4)})=0.000335$

Day 0 $\quad$ Day 2 $\quad$ 6

$\Pr(T>a+b \mid T>a)=\Pr(T>b)$ $\quad$ Memoryless

Gamma Distribution

Gamma function $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx$

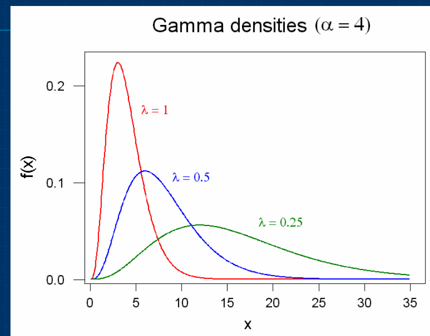
$\Gamma(1) = 1$ $\quad \Gamma(\alpha) = (\alpha-1)\times\Gamma(\alpha-1)$ $\quad \Gamma(n) = (n-1)!$

Gamma distribution $X \sim \Gamma(\alpha, \lambda)$

$$f(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x} \quad , x > 0$$

$$E(X) = \frac{\alpha}{\lambda} \qquad Var(X) = \frac{\alpha}{\lambda^2}$$

Gamma Distribution

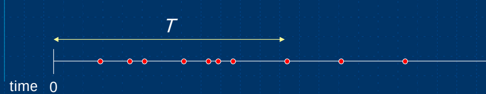


Gamma Distribution

Example : Average 20 phone calls per hour

Assume Poisson process

time unit = hour $\quad \lambda$ = 20



T = waiting time until the 8th call $\quad T \sim \Gamma(8, 20)$

$$E(T) = \frac{8}{20} = 0.4 \text{ hour} = 24 \text{ mins}$$

$$\Pr(T > 0.8) = 1 - \Pr(T \le 0.8) = 1 - \int_0^{0.8} \frac{(20)^8}{\Gamma(8)} x^7 e^{-20x} dx = ?$$

Chi-Square Distribution

$$X \sim \Gamma(\alpha, \lambda) \Leftrightarrow aX \sim \Gamma\left(\alpha, \frac{\lambda}{a}\right)$$

$$X \sim \Gamma(\alpha, \lambda) \Leftrightarrow 2\lambda X \sim \chi^2_{2\alpha}$$

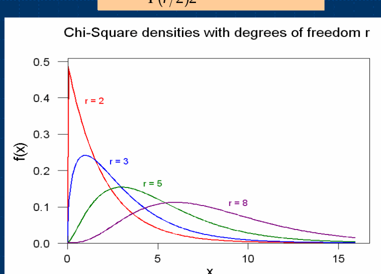
$$\chi^2_r \equiv \Gamma\left(\frac{r}{2}, \frac{1}{2}\right)$$

Example : $T \sim \Gamma(8, 20) \Leftrightarrow 40T \sim \Gamma\left(8, \frac{1}{2}\right) \equiv \chi^2_{16}$

$$\Pr(T \le 0.8) = \Pr(40T \le 32) = \Pr(\chi^2_{16} \le 32) = 0.99$$

Chi-Square Distribution

$$f(x) = \frac{1}{\Gamma(r/2) 2^{r/2}} x^{\frac{r}{2}-1} e^{-\frac{x}{2}} \quad , x > 0$$



$$E(X) = r \qquad Var(X) = 2r$$

Chi-Square Distribution Table

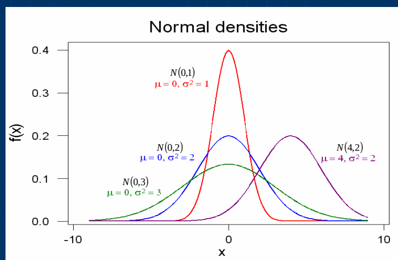
	$P(X \le x)$							
	0.010	0.025	0.050	0.100	0.900	0.950	0.975	0.990
r	$\chi^2_{0.99}(r)$	$\chi^2_{0.975}(r)$	$\chi^2_{0.95}(r)$	$\chi^2_{0.90}(r)$	$\chi^2_{0.10}(r)$	$\chi^2_{0.05}(r)$	$\chi^2_{0.025}(r)$	$\chi^2_{0.01}(r)$
1	0.000	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.34
4	0.297	0.484	0.711	1.064	7.779	9.488	11.14	13.28
5	0.554	0.831	1.145	1.610	9.236	11.07	12.83	15.09
6	0.872	1.237	1.635	2.204	10.64	12.59	14.45	16.81
7	1.239	1.690	2.167	2.833	12.02	14.07	16.01	18.48
8	1.646	2.180	2.733	3.490	13.36	15.51	17.54	20.09
9	2.088	2.700	3.325	4.168	14.68	16.92	19.02	21.67
10	2.558	3.247	3.940	4.865	15.99	18.31	20.48	23.21
11	3.053	3.816	4.575	5.578	17.28	19.68	21.92	24.72
12	3.571	4.404	5.226	6.304	18.55	21.03	23.34	26.22
13	4.107	5.009	5.892	7.042	19.81	22.36	24.74	27.69
14	4.660	5.629	6.571	7.790	21.06	23.68	26.12	29.14
15	5.229	6.262	7.261	8.547	22.31	25.00	27.49	30.58
16	5.812	6.908	7.962	9.312	23.54	26.30	28.84	32.00
17	6.408	7.564	8.672	10.08	24.77	27.59	30.19	33.41
18	7.015	8.231	9.390	10.86	25.99	28.87	31.53	34.80
19	7.633	8.907	10.12	11.65	27.20	30.14	32.85	36.19
20	8.260	9.591	10.85	12.44	28.41	31.41	34.17	37.57

$X \sim \chi^2_8$

$\Pr(X \le 15.51) = 0.95$

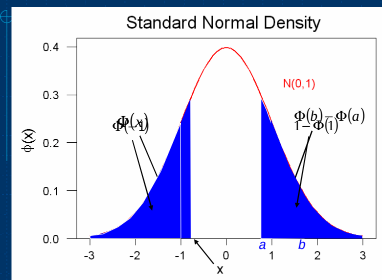
Normal Distribution

$$f(x)=\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}\quad,-\infty<x<\infty$$



$X\sim N(\mu,\sigma^2)$

Standard Normal Distribution



$X\sim N(0,1)$

$\Phi(x)=\Pr(X\le x)$

$\Pr(a\le X\le b)$
$=\Phi(b)-\Phi(a)$

$\Phi(-x)=1-\Phi(x)$

Standard Normal Distribution Table

$$P(Z\le z)=\Phi(z)=\int_{-\infty}^{z}\frac{1}{\sqrt{2\pi}}e^{-w^2/2}\,dw$$

$[\Phi(-z)=1-\Phi(z)]$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7703	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990

$X\sim N(0,1)$

$\Phi(-\Pr(X\le c)=\Phi(c)=0.9798\ 264$

$c=2.05$

Interpolation

$\Phi(0.917)\approx 0.8204$

Normal Distribution $X\sim N(\mu,\sigma^2)$

$aX+b\sim N(a\mu+b,a^2\sigma^2)$

$\frac{X-\mu}{\sigma}\sim N(0,1)$ Standard/Normal Score

$$\Pr(a\le X\le b)=\Pr\left(\frac{a-\mu}{\sigma}\le\frac{X-\mu}{\sigma}\le\frac{b-\mu}{\sigma}\right)$$
$$=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)$$

$\Pr(\mu-\sigma\le X\le\mu+\sigma)=\Phi(1)-\Phi(-1)=0.682=68.2\%$

$\Pr(\mu-2\sigma\le X\le\mu+2\sigma)=\Phi(2)-\Phi(-2)=0.944=94.4\%$

Normal Distribution

Example : Overall scores approximately normal with mean 74.3 and sd 11.8

$X\sim N(74.3,11.8^2)$

% of higher than 85 $\approx\Pr(X>85)=1-\Phi\left(\frac{85-74.3}{11.8}\right)$

$=1-\Phi(0.907)=1-0.8178=18.22\%$

95% can pass $\Rightarrow$ passing mark = ?

$\Pr(X\ge c)=0.95\Rightarrow 1-\Phi\left(\frac{c-74.3}{11.8}\right)=0.95$

$\Rightarrow\frac{c-74.3}{11.8}=-1.645\Rightarrow c=54.889\approx 55$