Topic 3: Random Variables and Important Probability Distributions

Optional Reading: Chapter 4

Xiner Zhou

Department of Statistics

University of California, Davis

- What is Random Variable and its Distribution?
- Discrete random variable
- Important discrete distribution
- Continuous random variable
- Important continuous distribution

A r.v. X is said to be continuous if:

X can take on any value in an interval in the real line.

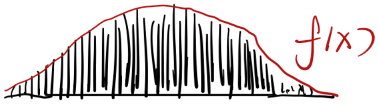
Probability Density Function (PDF) is the probability distribution for

Continuous random variables

f(x) with domain (a,b) that satisfies two properties:

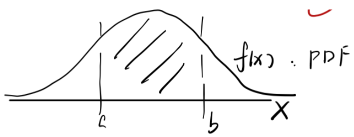
- $f(x) \geq 0$
- $\int_a^b f(x) = 1$

Discrete r.v.: PMF tells us the probability at each possible value



Think continuous r.r. as we increase possible values that X can take. PDF tells us relative frequency at all possible values.

Calculating probability of events using PDF



$$P(a < X < b) = \int_a^b f(x)dx$$

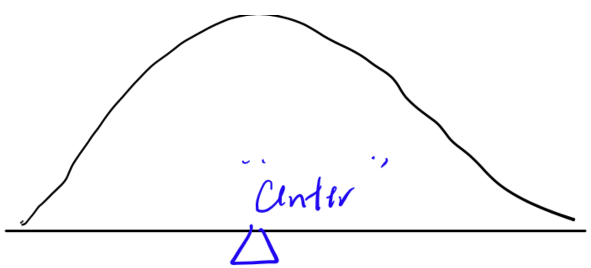
Probability is the area under the curve f(x) for x lies between a and b

Expectation / Expected Value/ Mean of a continuous r.v.:

$$\mu = E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

Values x PDF at x

Expectation of X is a weighted sum of possible values that X can take, weight being the probability that r.v. X =x



Variance of a continuous r.v. X is:

$$\sigma^2 = Var(X) = E(X - EX)^2 = \int_{-\infty}^{\infty} (x - EX)^2 f(x) dx$$

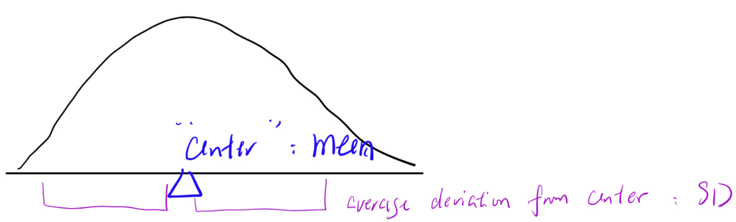
Variability/dispersion: How far away X is from its center, on average

Weighted sum of squared distance between each possible value x to the center mean E(X),
Weight being the probability of X=x

The square root of variance is called standard deviation (SD),

which measures variability in the original unit.

$$SD(X) = \sqrt{\mathrm{Var}(X)}$$



Among continuous distributions, Normal distribution is the most famous one.

It's extremely widely used because:

1. Many variables are nearly normal, because most measurement has large number of influencing factors, when those factors acting together, their additive effects often make the measurement we are interested in close to a Normal curve

2. One of the most important theorem — Central Limit Theorem — says that, the average of a large number of r.v.s has an approximately Normal distribution, regardless of what the actual distribution the random variables have.

   Meaning: averages are basically following Normal distribution.

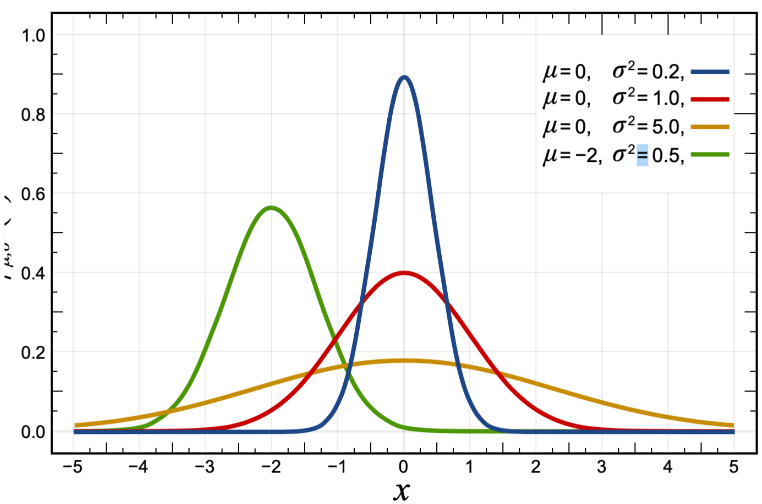
   Averages are what we care about most of times!

A continuous r.v. has a bell-shaped probability distribution or bell curve, is known as a normal random variable and its probability distribution is called a normal distribution.

Normal distribution has two parameters:

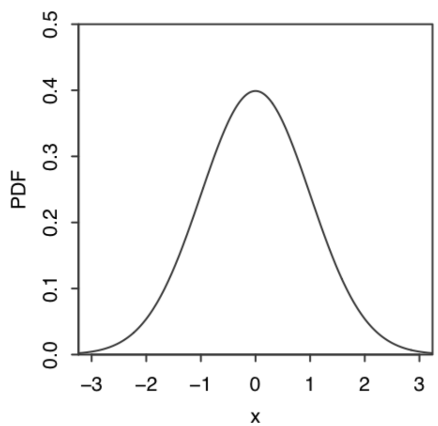
- Mean μ
- Standard deviation σ

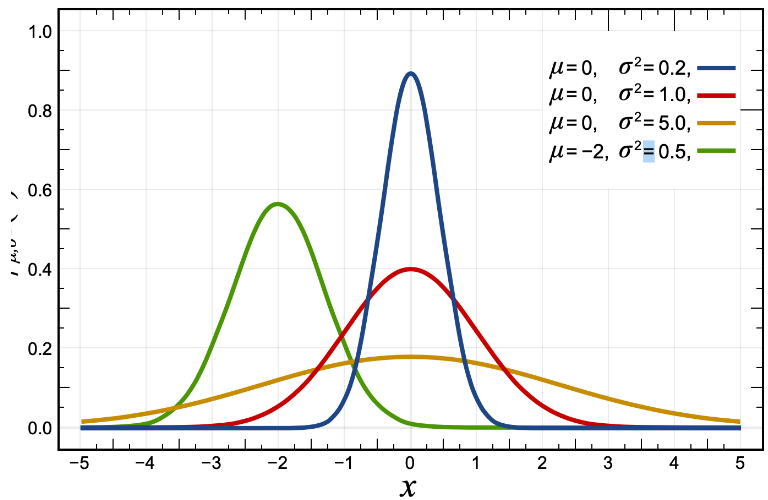
We denote $X \sim N(\mu, \sigma^2)$



The standard normal distribution is a normal distribution with $\mu = 0$ and $\sigma = 1$

We write $Z \sim N(0,1)$



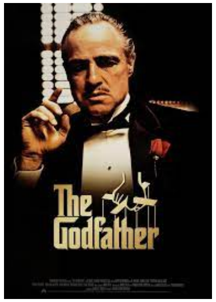


Symmetric around 0, bell-shaped curve

The Godfather:

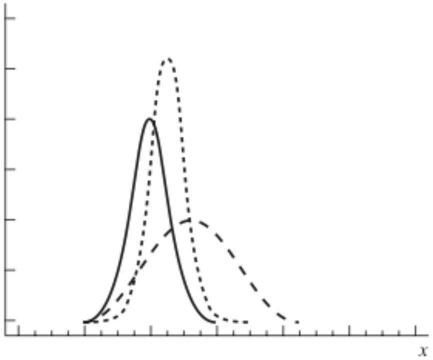
Standard normal distribution is just one member of the normal distribution family (Corleone family).

“There is nothing can’t be solved if you went to the Godfather for help.”



The general normal distribution is governed by:

- Mean μ — center
- Standard deviation σ — spread



Think of general normal distributions as a shift + a stretch upon standard normal.

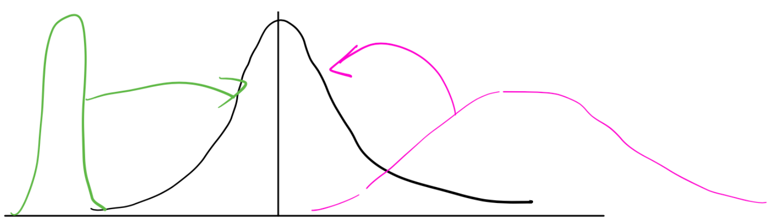
Z-score: describe how far X is away from its mean, in unit of its standard deviation

- If $X \sim N(\mu, \sigma^2)$, then $Z = \dfrac{X - \mu}{\sigma}$

You can turn any normal distribution into a standard one!

- If $Z \sim N(0,1)$, then $X = \mu + \sigma Z \sim N(\mu, \sigma^2)$

You can turn a standard one into any normal distribution!



Why do this? We know everything about the Godfather = Z!

To find normal probabilities, say P(X<a), one usually coverts X to its z-score, and uses standard normal table to find corresponding probabilities.

Steps for Finding a Probability Corresponding to a Normal Random Variable:

1. Sketch the normal distribution and indicate the mean of the random variable X. Then shade the area corresponding to the probability you want to find.

2. Convert the boundaries of the shaded area from X values to standard normal random variable z-score values by using the formula,

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

Show the z values under the corresponding X values on your sketch.

3. Use the standard normal table to find the areas corresponding to the z values. If necessary, use the symmetry of the normal distribution to find areas corresponding to negative z values and the fact that the total area on each side of the mean equals 0.5. to convert the areas from the table to the nrobabilities of the event you have selected

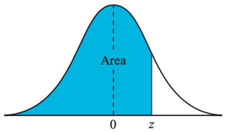


TABLE 3 **Areas under the Normal Curve**

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	.0017	.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	.0722	.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

TABLE 3 *(continued)*

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.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	.5 871	.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

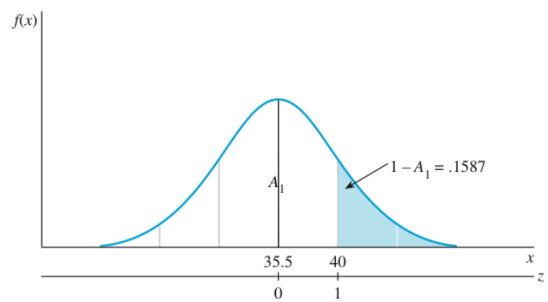
$p(z > 0)$, $p(z \leq -1.5)$, $p(z \geq 1.5)$, $p(-1.5 < z < 1.5)$, $p(-2.4 < z \leq 3)$, $p(z > 3.5)$

Studies show that gasoline use for compact cars sold in the United States is normally distributed, with a mean of 35.5 miles per gallon (mpg) and a standard deviation of $4.5 mpg. What percentage of compacts get ~~$~~40 mpg or more?

$$\boxed{X \sim N(35.5,\ 4.5^2)}$$

$$z = \frac{x - \mu}{\sigma} = \frac{40 - 35.5}{4.5} = 1.0$$

$$P(x \geq 40) = 1 - P(z < 1) = 1 - .8413 = .1587$$



$$P(X \geq 40)$$

$$= P\Big(\underbrace{\frac{X - 35.5}{4.5}}_{Z} \geq \frac{40 - 35.5}{4.5}\Big)$$

$$= P(Z \geq 1)$$

$$= 1 - P(Z < 1)$$

$$= 1 - 0.8413$$

$$= 0.1587$$

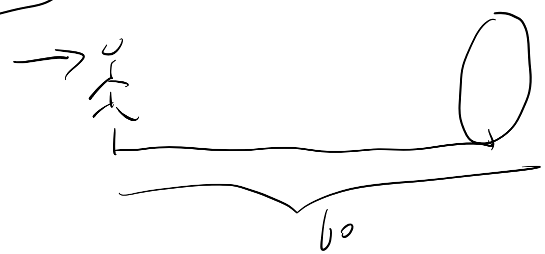
For a car traveling 30 miles per hour (mph), the distance required to brake to a stop is normally distributed with a mean of 50 feet and a standard deviation of 8 feet. Suppose you are traveling 30 mph in a residential area and a car moves abruptly into your path at a distance of 60 feet.

a. If you apply your brakes, what is the probability that you will brake to a stop within 40 feet or less? Within 50 feet or less?

b. If the only way to avoid a collision is to brake to a stop, what is the probability that you will avoid the collision?

$X \le 60$

b. $P(X \le 60)$



X = distance ... $\sim N(50, 8^2)$

a. $P(X \le 40) = P\left(\underbrace{\frac{X-50}{8}}_{Z} \le \underbrace{\frac{40-50}{8}}_{-10/8 = -1.25}\right)$

$= P(Z \le -1.25)$

$= 0.1056$

Once nice thing about Normal distribution is: easy to calculate probabilities of practical relevance based on the “68-95-99.7%” rule.

68-95-99.7% rule:

1. If $X \sim N\left(\mu, \sigma^2\right)$, then

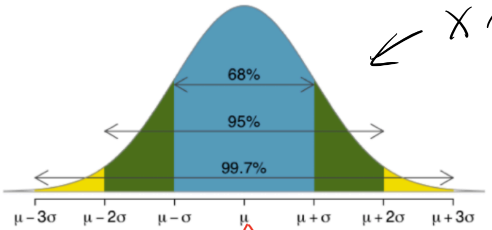
- $P(|\cancel{X} - \mu| < \sigma) \approx 0.68$
- $P(|\cancel{X} - \mu| < 2\sigma) \approx 0.95$
- $P(|\cancel{X} - \mu| < 3\sigma) \approx 0.997$

About 95% of the time, a normal data will fall within + or - 2 standard deviation away from its center

2. If $Z \sim N(0,1)$, then

- $P(|Z| < 1) \approx 0.68$
- $P(|Z| < 2\ ) \approx 0.95$
- $P(|Z| < 3) \approx 0.997$

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



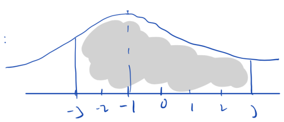
μ: center

σ: average distance from center

Let $X \sim N(-1,4)$, what is P(|X|<3) approximately?

$= \sigma^2$

Solution :



standardization $Z = \frac{X-(-1)}{2} = \frac{X+1}{2} \sim N(0,1)$

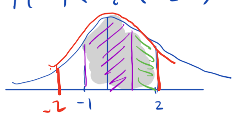
$P(\underbrace{|X|<3}_{\text{try to express this event in terms of } Z})$

$= P(-3 < X < 3)$

$= P\left(\frac{-3+1}{2} < \frac{X+1}{2} < \frac{3+1}{2}\right)$

$= P\left(-1 < \underbrace{\frac{X+1}{2}}_{Z} < 2\right)$

$= P(-1 < Z < 2)$



68-95-99.7% rule : $P(-1 < Z < 1) \approx 0.68$

$P(-2 < Z < 2) \approx 0.95$

$P(\;) \approx 0.68$

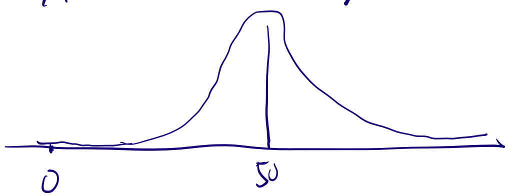
$P(\;) \approx \frac{0.95-0.68}{2} = \frac{0.27}{2}$

$\approx 0.68 + \frac{0.27}{2} = 0.815$

A normal random variable X has mean 50 and standard deviation 15.

Would it be unusual to observe the value x=0? Explain your answer.

$X \sim N(50, 15^2)$

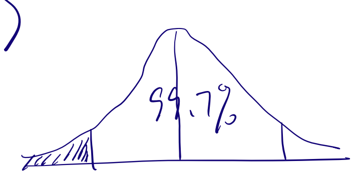


$$P(X \le 0) = P\left(\frac{X-50}{15} \le \frac{0-50}{15}\right)$$

$$= P(Z \le -3.xxx)$$

$$\le \frac{1-0.997}{2}$$

$$\approx 0.00xxx$$



$$\frac{1-99.7\%}{2}.$$