Math 1070-2: Spring 2008

Lecture 6

Davar Khoshnevisan

Department of Mathematics
University of Utah
http://www.math.utah.edu/~davar

February 20, 2008

Probability distributions

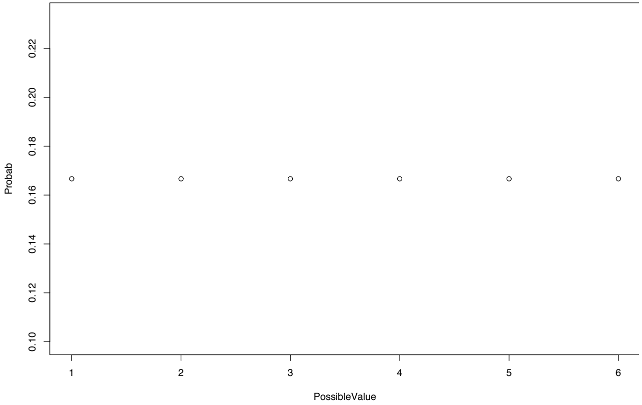
- Recall that a probability distribution is a table of possible values versus their probabilities
- Example:

value	1	2	3	4	5	6
probability	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

- The shape is flat.

Rolling one die

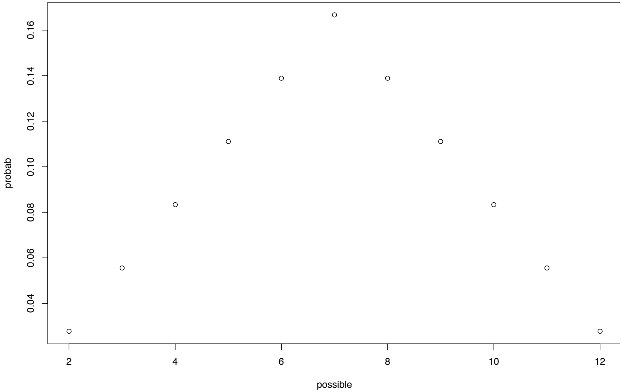
- Sample space = $\{1, 2, \ldots, 6\}$ (all equally likely)
- Possible values = 1, 2, 3, 4, 5, 6
- Resp. probab. = $\frac{1}{6}$ each



-

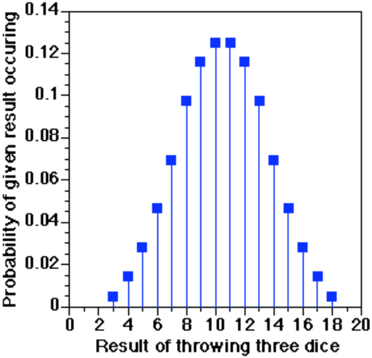
Rolling two dice

- Sample space = $\{(1,1),\ldots,(6,6)\}$ (all equally likely)
- Possible values = 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12
- Resp. probab. = $\frac{1}{36}, \frac{2}{36}, \frac{3}{36}, \frac{4}{36}, \frac{5}{36}, \frac{6}{36}, \frac{5}{36}, \frac{4}{36}, \frac{3}{36}, \frac{2}{36}, \frac{1}{36}$



-

Rolling three dice



The mean/SD of a probability distribution

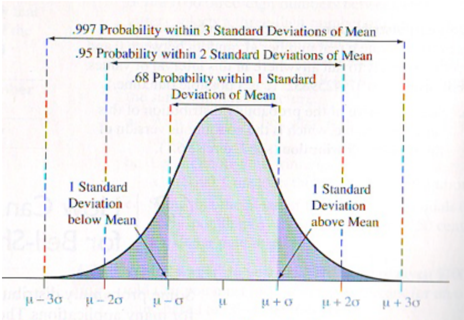
- Recall: The mean is $\mu = \sum xP(x)$
- E.g., one die: the expected number of dots is

$$\left(1 \times \frac{1}{6}\right) + \cdots + \left(6 \times \frac{1}{6}\right) = 3.5$$

- There is also a notion of SD (σ) which measures the average deviation from μ in the population

The normal distribution

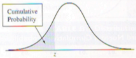
- Two parameters $\mu =$ the mean; $\sigma =$ st. dev.



-
- How? Why?

The standard normal table (page 1)

Appendix A



Cumulative probability for z is the area under the standard normal curve to the left of z

z	.00
−5.0	.000000287
−4.5	.00000340
−4.0	.0000317
−3.5	.000233

TABLE A Standard Normal Cumulative Probabilities

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

The general normal table

- General fact: If you change x to SUs then you do not alter the probabilities
- I.e., the probability of being z SDs above [or below] the mean does not depend on μ or σ
- Only true for normal distributions!!!!!
- "Formula":

$$\text{height at } x = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{z^2}{2}}$$

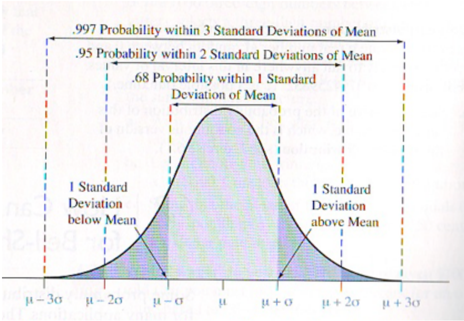
A blackboard example (SATs)

- SAT scores $\approx$ normal with $\mu = 500$ and SD$= 100$
- If your score is $x = 650$ then

$$z = \frac{x - \mu}{\sigma} = \frac{650 - 500}{100} = 1.5.$$

- What is the percentage of scores less than yours?
  (this $\times 100\%$ is your score's percentile)
- draw a picture!
- Ans = $(1 - 0.0668) \times 100\% \approx 93.3\%$

Basic normal probab.s

- 

- Blackboard computations (draw pictures!)

The binomial distribution

- n independent trials
- each leads to two possible outcomes (success/failure, man/woman, smoker/nonsmoker, ...)
- Each trial has the same chance p of leading to a success
- Probab. of getting x successes is exactly:

$$\frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \quad \text{for } x = 0, 1, \ldots, n$$

- $z! = z$ factorial $= z \times (z-1) \times (z-2) \times \cdots \times 2 \times 1$
- $0! = 1$

The binomial distribution (Example)

- Roughly half of a large population is men $p = \frac{1}{2} = 0.5$
- Sample 10 people independently (n=10)
- Find probab. of no women in the sample
- Probab. of getting x women is exactly:

$$\frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} \quad \text{for } x = 0, 1, \ldots, n$$

- Set $n = 10$, $p = 0.5$, and $x = 0$ to find that

$$\text{probab. no women} = \frac{10!}{0! \times (10-0)!} 0.5^0 (1-0.5)^{10-0} = 0.001$$

Facts about binomials

- $\mu = np$
- $\sigma = \sqrt{np(1-p)}$
- In the previous example $(n = 10, p = 0.5)$ of women vs men,

$$\mu = 10 \times 0.5 = 5 \text{ women} \quad \text{and} \quad \sigma = \sqrt{10 \times 0.5 \times (1 - 0.5)} \approx 1.5$$

- Deep fact: If n is large then binomial (n, p) probab.s are close to those of a normal with $\mu = np$ and $\sigma = \sqrt{np(1-p)}$

Example (racial profiling)

- 1990s: US Justice Dept, ACLU, etc. studied possible abuse by Philadelphia PD's treatment of minorities
- Results of 262 ($n = 262$) police-car stops during a certain week in 1997:
  207 (79%) of the drivers were African American
- Is this unusual?
- Suppose the percentage of African Americans in Philly in 1997 $\approx$ that in the US (42.2%; $p = 0.422$)
- If no profiling, then the no. of African Amercians in the sample is binomial with $n = 262$ and $p = 0.422$ (Why? Model?)

Example (racial profiling; continued)

- $\mu = 262 \times 0.422 \approx 110.563$
- $\sigma = \sqrt{262 \times 0.422(1 - 0.422)} \approx 7.99$
- If $x = 207$ then $z = (x - 110.563)/7.99 \approx 12$
- Interpret using a normal table
- One possible limitation of this analysis: Were 42.2% of all possible stops African Americans?