Commonly Used Distributions and Parameter Estimation

Berlin Chen
Department of Computer Science & Information Engineering
National Taiwan Normal University

Reference:
1. W. Navidi. *Statistics for Engineering and Scientists*. Chapter 4 & Teaching Material

How to Estimate Population (Distribution) Parameters ?

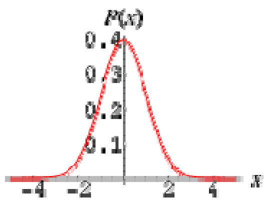
Population

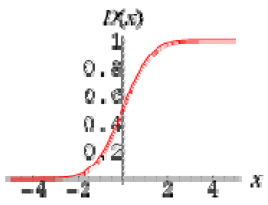
Sample

Inference

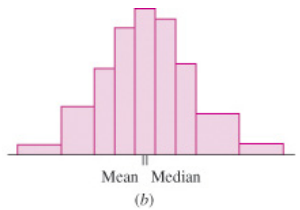
Statistics

Parameters





- Unbiased Estimate ?
- Estimation Uncertainty ?



The Bernoulli Distribution

- We use the Bernoulli distribution when we have an experiment which can result in one of two outcomes
  - One outcome is labeled "success," and the other outcome is labeled "failure"
  - The probability of a success is denoted by p. The probability of a failure is then $1 - p$

- Such a trial is called a Bernoulli trial with success probability p

Examples

1. The simplest Bernoulli trial is the toss of a coin. The two outcomes are heads and tails. If we define heads to be the success outcome, then p is the probability that the coin comes up heads. For a fair coin, $p = ½$

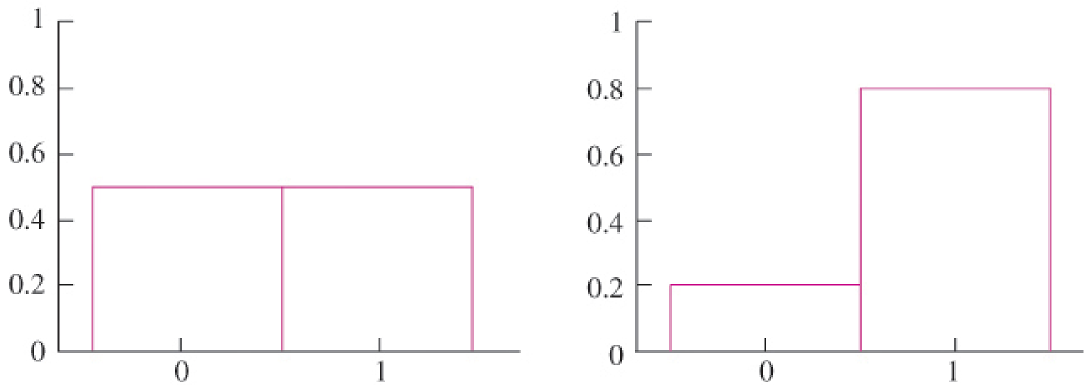
2. Another Bernoulli trial is a selection of a component from a population of components, some of which are defective. If we define "success" to be a defective component, then p is the proportion of defective components in the population

X ~ Bernoulli(p)

- For any Bernoulli trial, we define a random variable X as follows:
  - If the experiment results in a success, then $X = 1$. Otherwise, $X = 0$. It follows that X is a discrete random variable, with probability mass function $p(x)$ defined by

$$p(0) = P(X = 0) = 1 - p$$
$$p(1) = P(X = 1) = p$$

$p(x) = 0$ for any value of x other than 0 or 1



(a) (b)

Mean and Variance of Bernoulli

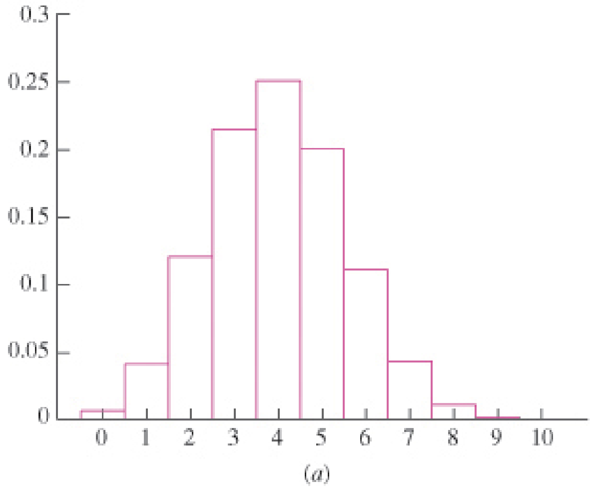
- If X ~ Bernoulli(p), then
  - $\mu_X = 0(1- p) + 1(p) = p$
  - $\sigma_X^2 = (0-p)^2(1-p) + (1-p)^2(p) = p(1-p)$

The Binomial Distribution

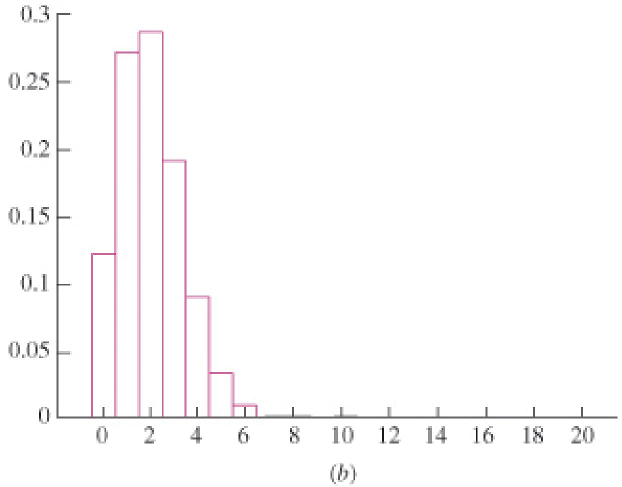
- If a total of n Bernoulli trials are conducted, and
  - The trials are independent
  - Each trial has the same success probability p
  - X is the number of successes in the n trials

Then X has the binomial distribution with parameters n and p, denoted $X \sim \text{Bin}(n,p)$

Probability Histogram



Bin(10, 0.4)



Bin(20, 0.1)

Another Use of the Binomial

- Assume that a finite population contains items of two types, successes and failures, and that a simple random sample is drawn from the population. Then if the sample size is no more than 5% of the population, the binomial distribution may be used to model the number of successes
  - Sample items can be therefore assumed to be independent of each other
  - Each sample item is a Bernoulli trial

pmf, Mean and Variance of Binomial

- If X ~ Bin(n, p), the probability mass function of X is

$$p(x) = P(X = x) = \begin{cases} \dfrac{n!}{x!(n-x)!} p^x (1-p)^{n-x}, x = 0, 1, \ldots, n \\ 0, \quad \text{otherwise} \end{cases}$$

$$\binom{n}{x} = \frac{n!}{x!(n-x)!}$$

  - Mean: $\mu_X = np$
  - Variance: $\sigma_X^2 = np(1-p)$

More on the Binomial

- Assume n independent Bernoulli trials are conducted
- Each trial has probability of success p
- Let $Y_1, \ldots, Y_n$ be defined as follows: $Y_i = 1$ if the i^{th} trial results in success, and $Y_i = 0$ otherwise (Each of the Y_i has the Bernoulli(p) distribution)
- Now, let X represent the number of successes among the n trials. So, $X = Y_1 + \ldots + Y_n$

⇨ This shows that a binomial random variable can be expressed as a sum of Bernoulli random variables

Estimate of *p*

- If $X \sim \mathrm{Bin}(n, p)$, then the sample proportion $\hat{p} = X / n$

$$\hat{p} = \frac{\text{number of successes}}{\text{number of trials}} = \frac{X}{n} \quad \left(= \frac{Y_1 + Y_2 + \cdots + Y_n}{n}\right)$$

  is used to estimate the success probability *p*

- Note:
  - Bias is the difference $\mu_{\hat{p}} - p$.
  - $\hat{p}$ is unbiased $\left(\mu_{\hat{p}} - p = 0\right)$
  - The uncertainty in $\hat{p}$ is

$$\sigma_{\hat{p}} = \sigma_{(Y_1 + Y_2 + \cdots + Y_n)/n} = \sqrt{\frac{p(1-p)}{n}}$$

  - In practice, when computing σ, we substitute $\hat{p}$ for *p*, since *p* is unknown

The Poisson Distribution

- One way to think of the Poisson distribution is as an approximation to the binomial distribution when n is large and p is small

- It is the case when n is large and p is small the mass function depends almost entirely on the mean np, very little on the specific values of n and p

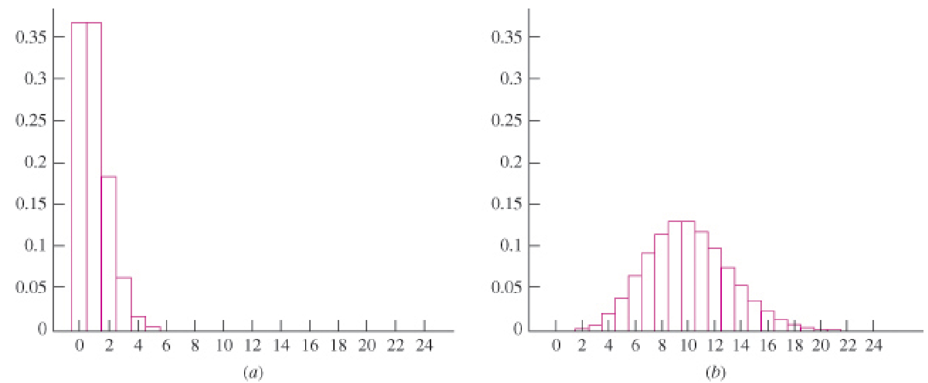
- We can therefore approximate the binomial mass function with a quantity $\lambda = np$; this λ is the parameter in the Poisson distribution

pmf, Mean and Variance of Poisson

- If X ~ Poisson(λ), the probability mass function of X is

$$p(x) = P(X = x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^{x}}{x!}, & \text{for } x = 0,\ 1,\ 2,\ \ldots \\ 0, & \text{otherwise} \end{cases}$$

Probability Histogram



Poisson(1) Poisson(10)

  - Mean: $\mu_X = \lambda$
  - Variance: $\sigma_X^2 = \lambda$

- Note: X must be a discrete random variable and λ must be a positive constant

Relationship between Binomial and Poisson

- The Poisson PMF with parameter λ is a good approximation for a binomial PMF with parameters n and p, provided that $\lambda = np$, n is very large and p is very small

$$\lim_{n\to\infty}\binom{n}{k}p^k(1-p)^{n-k}$$

$$=\lim_{n\to\infty}\frac{n!}{(n-k)!k!}p^k(1-p)^{n-k} \quad (\because \lambda = np \Rightarrow p = \frac{\lambda}{n})$$

$$=\lim_{n\to\infty}\frac{n(n-1)\cdots(n-k+1)}{k!}\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$=\lim_{n\to\infty}\frac{\lambda^k}{k!}\frac{n(n-1)\cdots(n-k+1)}{n^k}\left(1-\frac{\lambda}{n}\right)^{n-k}$$

$$=\lim_{n\to\infty}\frac{\lambda^k}{k!}\left(\frac{n}{n}\right)\left(\frac{n-1}{n}\right)\cdots\left(\frac{n-k+1}{n}\right)\left(1-\frac{\lambda}{n}\right)^{-k}\left(1-\frac{\lambda}{n}\right)^{n} \qquad (\because \lim_{n\to\infty}\left(1+\frac{x}{n}\right)^n = e^x)$$

$$=\lim_{n\to\infty}\frac{\lambda^k}{k!}e^{-\lambda}$$

TABLE 4.1 An example of the Poisson approximation to the binomial probability mass function*

x	$P(X = x)$, $X \sim$ Bin (10,000, 0.0002)	$P(Y = x)$, $Y \sim$ Bin (5000, 0.0004)	Poisson Approximation Poisson (2)
0	0.135308215	0.135281146	0.135335283
1	0.270670565	0.270670559	0.270670566
2	0.270697637	0.270724715	0.270670566
3	0.180465092	0.180483143	0.180447044
4	0.090223521	0.090223516	0.090223522
5	0.036082189	0.036074965	0.036089409
6	0.012023787	0.012017770	0.012029803
7	0.003433993	0.003430901	0.003437087
8	0.000858069	0.000856867	0.000859272
9	0.000190568	0.000190186	0.000190949

*When n is large and p is small, the Bin(n, p) probability mass function is well approximated by the Poisson (λ) probability mass function (Equation 4.9), with $\lambda = np$. Here $X \sim$ Bin(10,000, 0.0002) and $Y \sim$ Bin(5000, 0.0004), so $\lambda = np = 2$, and the Poisson approximation is Poisson(2).

Poisson Distribution to Estimate Rate

- Let λ denote the mean number of events that occur in one unit of time or space. Let X denote the number of events that are observed to occur in t units of time or space
- If $X \sim$ Poisson(λt), we estimate λ with $\hat{\lambda} = \frac{X}{t}$
- Note:
  - $\hat{\lambda}$ is unbiased ($\mu_{\hat{\lambda}} = \mathbf{E}[\hat{\lambda}] = \mathbf{E}\left[\frac{X}{t}\right] = \frac{1}{t}\mathbf{E}[X] = \frac{1}{t} \cdot \lambda \cdot t = \lambda$)
  - The uncertainty in $\hat{\lambda}$ is $\sigma_{\hat{\lambda}} = \sigma_{\frac{X}{t}} = \sqrt{\frac{1}{t^2}\sigma_X^2} = \sqrt{\frac{1}{t^2}\lambda t} = \sqrt{\frac{\lambda}{t}}$
  - In practice, we substitute $\hat{\lambda}$ for λ, since λ is unknown

Some Other Discrete Distributions

- Consider a finite population containing two types of items, which may be called successes and failures
  - A simple random sample is drawn from the population
  - Each item sampled constitutes a Bernoulli trial
  - As each item is selected, the probability of successes in the remaining population decreases or increases, depending on whether the sampled item was a success or a failure
  - For this reason the trials are not independent, so the number of successes in the sample does not follow a binomial distribution

- The distribution that properly describes the number of successes is the hypergeometric distribution

pmf of Hypergeometric

- Assume a finite population contains N items, of which R are classified as successes and $N - R$ are classified as failures
  - Assume that n items are sampled from this population, and let X represent the number of successes in the sample
  - Then X has a hypergeometric distribution with parameters N, R, and n, which can be denoted $X \sim \mathrm{H}(N,R,n)$. The probability mass function of X is

$$p(x) = P(X = x) = \begin{cases} \dfrac{\dbinom{R}{x}\dbinom{N-R}{n-x}}{\dbinom{N}{n}}, & \text{if } \max(0, R+n-N) \le x \le \min(n, R) \\ 0, & \text{otherwise} \end{cases}$$

Mean and Variance of Hypergeometric

- If $X \sim$ H(N, R, n), then
  - Mean of X: $\mu_X = \dfrac{nR}{N}$
  - Variance of X: $\sigma_X^2 = n\left(\dfrac{R}{N}\right)\left(1-\dfrac{R}{N}\right)\left(\dfrac{N-n}{N-1}\right)$

Geometric Distribution

- Assume that a sequence of independent Bernoulli trials is conducted, each with the same probability of success, p
- Let X represent the number of trials up to and including the first success
  - Then X is a discrete random variable, which is said to have the geometric distribution with parameter p.
  - We write $X \sim \text{Geom}(p)$.

pmf, Mean and Variance of Geometric

- If X ~ Geom(p), then
  - The pmf of X is $p(x) = P(X = x) = \begin{cases} p(1-p)^{x-1}, \ x = 1, 2, \ldots \\ 0, \quad \text{otherwise} \end{cases}$
  - The mean of X is $\mu_X = \dfrac{1}{p}$
  - The variance of X is $\sigma_X^2 = \dfrac{1-p}{p^2}$

Negative Binomial Distribution

- The negative binomial distribution is an extension of the geometric distribution. Let r be a positive integer. Assume that independent Bernoulli trials, each with success probability p, are conducted, and let X denote the number of trials up to and including the r^{th} success
  - Then X has the negative binomial distribution with parameters r and p. We write $X \sim \text{NB}(r,p)$

- Note: If $X \sim \text{NB}(r,p)$, then $X = Y_1 + \ldots + Y_r$ where $Y_1, \ldots, Y_r$ are independent random variables, each with Geom(p) distribution

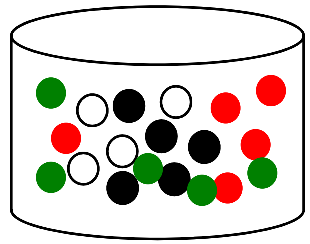
$$\underbrace{0,\ldots,1,}_{y_1}\ \underbrace{0,\ldots,1,}_{y_2}\ \underbrace{0,\ldots 1,}_{}\ \ldots\ldots,\underbrace{0,\ldots,1}_{y_r}$$

x trials in total

pmf, Mean and Variance of Negative Binomial

- If X ~ NB(r,p), then
  - The pmf of X is $p(x)=P(X=x)=\begin{cases}\binom{x-1}{r-1}p^r(1-p)^{x-r}, & x=r,r+1,\ldots\\ 0, & \text{otherwise}\end{cases}$
  - The mean of X is $\mu_X=\frac{r}{p}$
  - The variance of X is $\sigma_X^2=\frac{r(1-p)}{p^2}$

Multinomial Distribution



- A Bernoulli trial is a process that results in one of two possible outcomes. A generalization of the Bernoulli trial is the multinomial trial, which is a process that can result in any of k outcomes, where $k \geq 2$. We denote the probabilities of the k outcomes by $p_1,\ldots,p_k$ $(p_1+\ldots+p_k=1)$

- Now assume that n independent multinomial trials are conducted each with k possible outcomes and with the same probabilities $p_1,\ldots,p_k$. Number the outcomes 1, 2, …, k. For each outcome i, let X_i denote the number of trials that result in that outcome. Then $X_1,\ldots,X_k$ are discrete random variables. The collection $X_1,\ldots,X_k$ is said to have the multinomial distribution with parameters $n, p_1,\ldots,p_k$. We write $X_1,\ldots,X_k \sim \mathrm{MN}(n, p_1,\ldots,p_k)$

pmf of Multinomial

- If $X_1,\ldots,X_k \sim \mathrm{MN}(n, p_1,\ldots,p_k)$, then the pmf of $X_1,\ldots,X_k$ is

$$p(x) = P(X = x) = \begin{cases} \dfrac{n!}{x_1!x_2!\cdots x_k!} p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}, \; x_i = 0,1,2,\ldots,n \\ \qquad\qquad\qquad\qquad\qquad \text{and } \sum x_i = n \\ 0, \quad \text{otherwise} \end{cases}$$

Can be viewed as a joint probability mass function of $X_1,\ldots,X_k$

- Note that if $X_1,\ldots,X_k \sim \mathrm{MN}(n, p_1,\ldots,p_k)$, then for each i, $X_i \sim \mathrm{Bin}(n, p_i)$

The Normal Distribution

- The normal distribution (also called the Gaussian distribution) is by far the most commonly used distribution in statistics. This distribution provides a good model for many, although not all, continuous populations

- The normal distribution is continuous rather than discrete. The mean of a normal population may have any value, and the variance may have any positive value

pmf, Mean and Variance of Normal

- The probability density function of a normal population with mean μ and variance σ^2 is given by

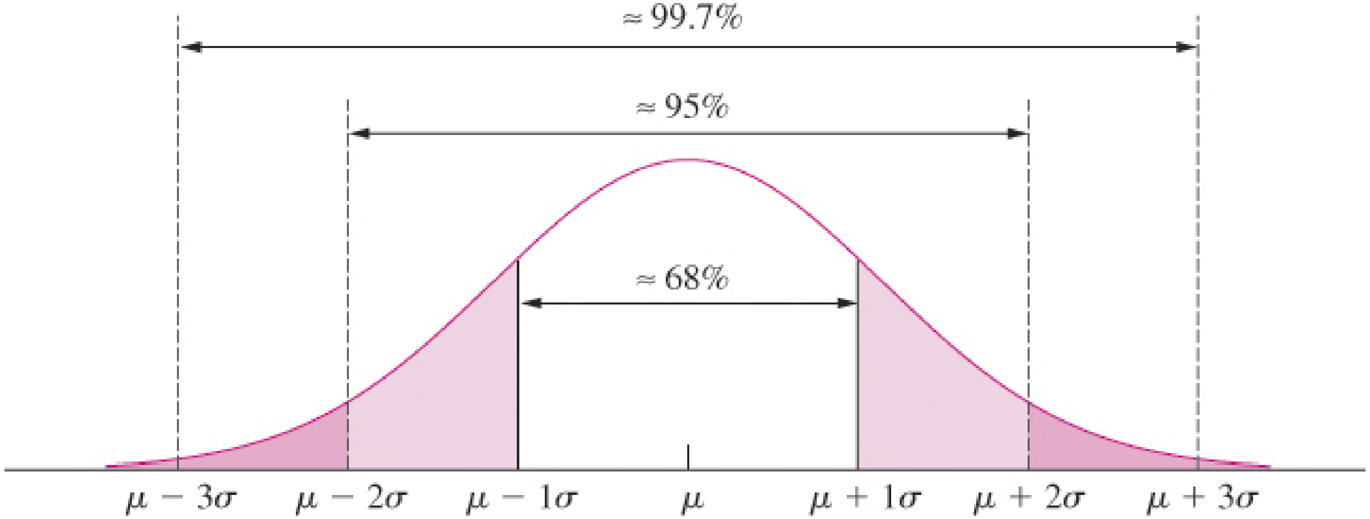
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}, \quad -\infty < x < \infty$$

- If $X \sim N(\mu, \sigma^2)$, then the mean and variance of X are given by

$$\mu_X = \mu$$

$$\sigma_X^2 = \sigma^2$$

68-95-99.7% Rule



- The above figure represents a plot of the normal probability density function with mean μ and standard deviation σ. Note that the curve is symmetric about μ, so that μ is the median as well as the mean. It is also the case for the normal population
  - About 68% of the population is in the interval $\mu \pm \sigma$
  - About 95% of the population is in the interval $\mu \pm 2\sigma$
  - About 99.7% of the population is in the interval $\mu \pm 3\sigma$

Standard Units

- The proportion of a normal population that is within a given number of standard deviations of the mean is the same for any normal population

- For this reason, when dealing with normal populations, we often convert from the units in which the population items were originally measured to standard units

- Standard units tell how many standard deviations an observation is from the population mean

Standard Normal Distribution

- In general, we convert to standard units by subtracting the mean and dividing by the standard deviation. Thus, if x is an item sampled from a normal population with mean μ and variance σ^2, the standard unit equivalent of x is the number z, where

$$z = (x - \mu)/\sigma$$

- The number z is sometimes called the "z-score" of x. The z-score is an item sampled from a normal population with mean 0 and standard deviation of 1. This normal distribution is called the standard normal distribution

Examples

1. Q: Aluminum sheets used to make beverage cans have thicknesses that are normally distributed with mean 10 and standard deviation 1.3. A particular sheet is 10.8 thousandths of an inch thick. Find the *z*-score:

   Ans.: $z = (10.8 - 10)/1.3 = 0.62$

2. Q: Use the same information as in 1. The thickness of a certain sheet has a z-score of -1.7. Find the thickness of the sheet in the original units of thousandths of inches:

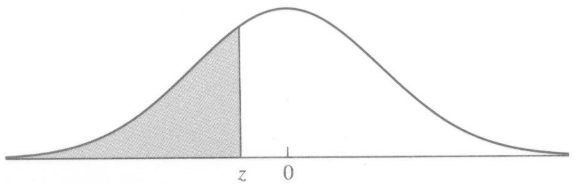
   Ans.: $-1.7 = (x - 10)/1.3 \quad x = -1.7(1.3) + 10 = 7.8$

Finding Areas Under the Normal Curve

- The proportion of a normal population that lies within a given interval is equal to the area under the normal probability density above that interval. This would suggest integrating the normal pdf; this integral have no closed form solution
- So, the areas under the curve are approximated numerically and are available in Table A.2 (Z-table). This table provides area under the curve for the standard normal density. We can convert any normal into a standard normal so that we can compute areas under the curve
  - The table gives the area in the left-hand tail of the curve
  - Other areas can be calculated by subtraction or by using the fact that the total area under the curve is 1

Z-Table (1/2)

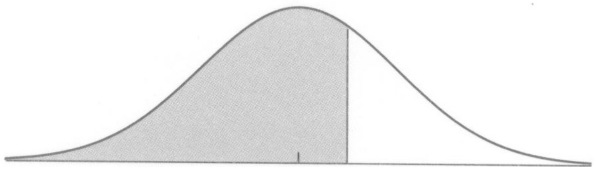
TABLE A.2 Cumulative normal distribution (z table)



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
−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

Z-Table (2/2)

TABLE A.2 Cumulative normal distribution (continued)

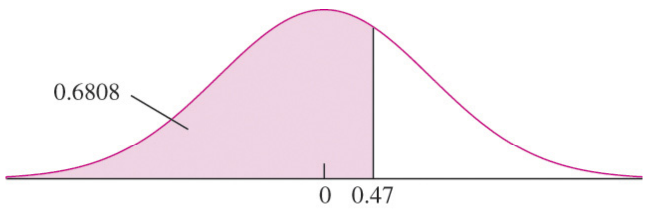


z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.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

Examples

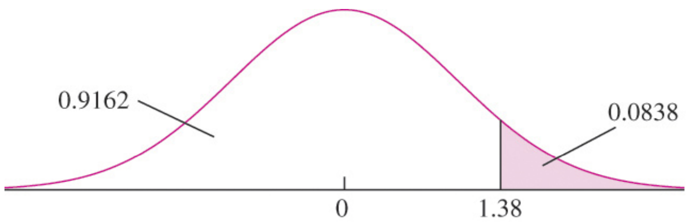
1. Q: Find the area under normal curve to the left of z = 0.47

   Ans.: From the z table, the area is 0.6808



2. Q: Find the area under the curve to the right of z = 1.38

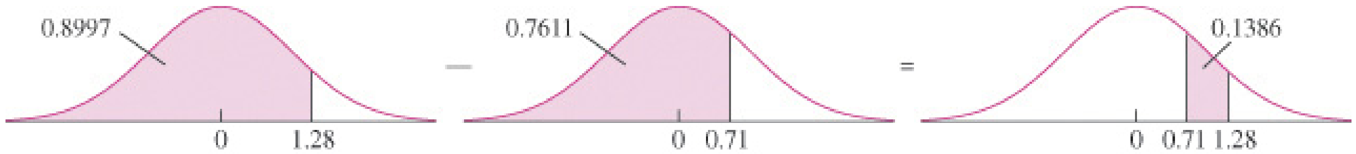
   Ans.: From the z table, the area to the left of 1.38 is 0.9162. Therefore the area to the right is 1 – 0.9162 = 0.0838



More Examples

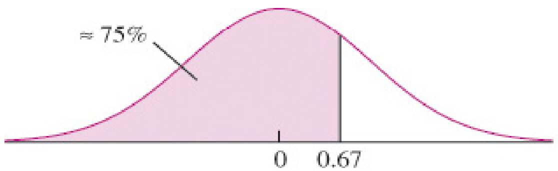
1. Q: Find the area under the normal curve between z = 0.71 and z = 1.28.

   Ans.: The area to the left of z = 1.28 is 0.8997. The area to the left of z = 0.71 is 0.7611. So the area between is 0.8997 – 0.7611 = 0.1386



2. Q: What z-score corresponds to the 75^{th} percentile of a normal curve?

   Ans.: To answer this question, we use the z table in reverse. We need to find the z-score for which 75% of the area of curve is to the left. From the body of the table, the closest area to 75% is 0.7486, corresponding to a z-score of 0.67

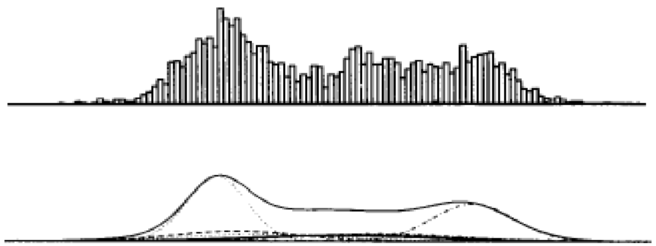


Linear Combinations of Independent Normal RVs

- The linear combinations of independent normal random variables are still normal random variables
  - Let $X_1 \sim N\left(\mu_1, \sigma_1^2\right), \ldots, X_n \sim N\left(\mu_n, \sigma_n^2\right)$ are independent, then

    $Y = c_1 X_1 + \cdots + c_n X_n$ is normal with
    - Mean $\mu_Y = c_1 \mu_1 + \cdots + c_n \mu_n$
    - Variance $\sigma_Y^2 = c_1^2 \sigma_1^2 + \cdots + c_n^2 \sigma_n^2$

- We have to distinguish the meaning of $Y = c_1 X_1 + \cdots + c_n X_n$ from that of $f_Y(y) = c_1 f_{X_1}(y) + \cdots + c_n f_{X_n}(y)$ $\sum_{i=1}^{n} c_i = 1$



Evaluating an Estimator : Bias and Variance (1/3)

- The mean square error of the estimator d can be further decomposed into two parts respectively composed of bias and variance

$$r(d,\theta) = E\left[(d-\theta)^2\right]$$

(Mean Squared Error, MSE-- mean of the squared error)

$$= E\left[(d - E[d] + E[d] - \theta)^2\right]$$

$$= E\left[(d - E[d])^2 + (E[d]-\theta)^2 + 2(d - E[d])(E[d]-\theta)\right]$$

$$= E\left[(d - E[d])^2\right] + E\left[\underbrace{(E[d]-\theta)}_{\text{constant}}{}^2\right] + 2E\left[(d - E[d])\underbrace{(E[d]-\theta)}_{\text{constant}}\right]$$

$$= E\left[(d - E[d])^2\right] + (E[d]-\theta)^2 + 2\underbrace{E\left[(d - E[d])\right]}_{0}(E[d]-\theta)$$

$$= \underbrace{E\left[(d - E[d])^2\right]}_{\text{variance}} + \underbrace{(E[d]-\theta)^2}_{\text{bias}^2}$$

Cf. Section 4.9

Evaluating an Estimator : Bias and Variance (2/3)

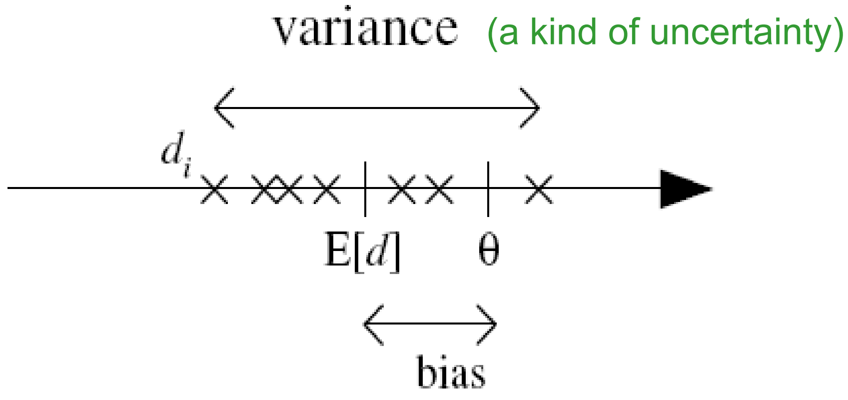
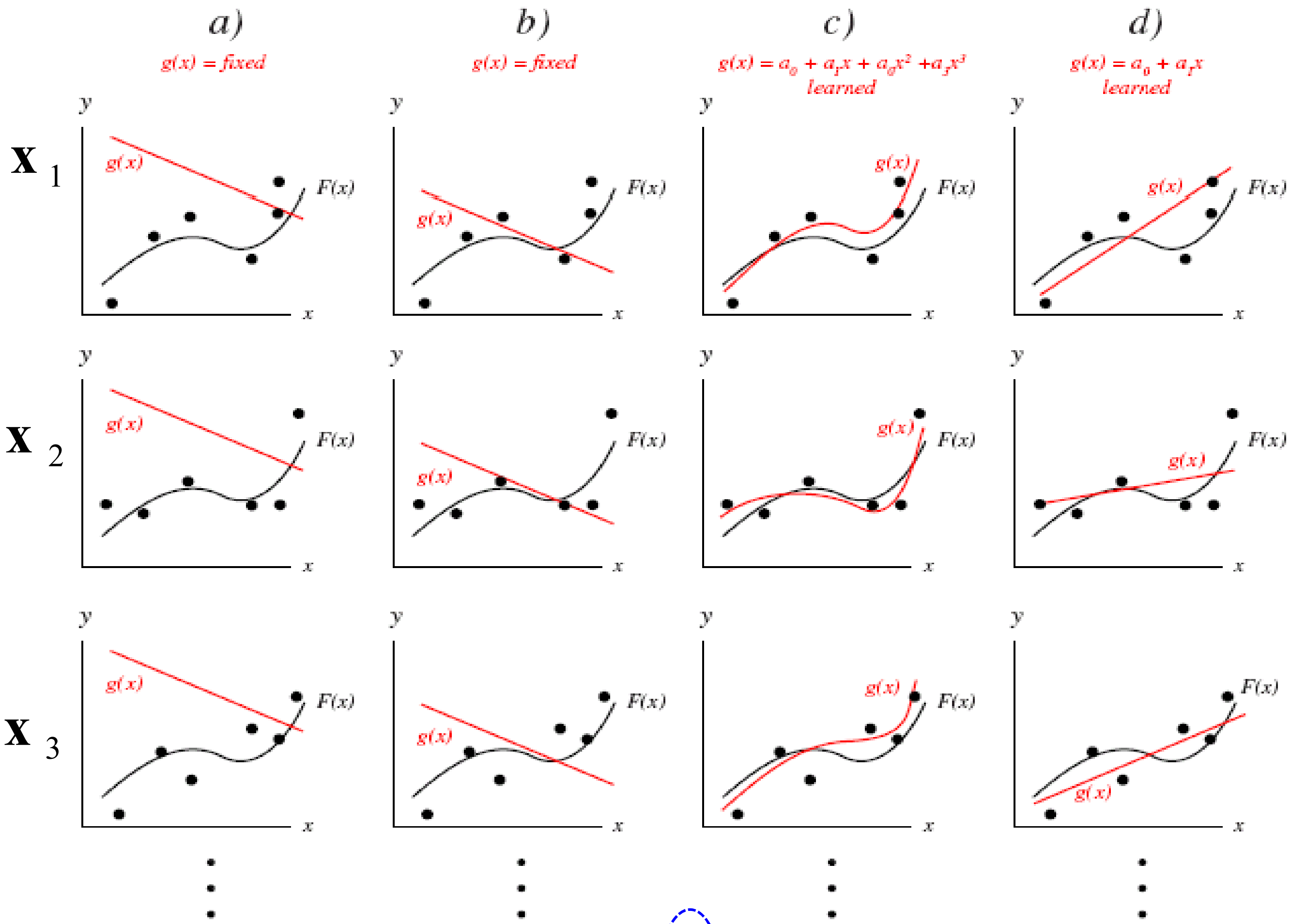


Figure 4.1: θ is the parameter to be estimated. d_i are several estimates (denoted by '$\times$') over different samples. Bias is the difference between the expected value of d and θ. Variance is how much d_i are scattered around the expected value. We would like both to be small.

Evaluating an Estimator : Bias and Variance (3/3)

- Bias and Variance: An Example



different samples for an unknown population

$$X \rightarrow (x, y)$$

$$y = F(x)$$

$$y' = F(x) + \varepsilon$$

error of measurement

Estimating the Parameters of Normal

- If $X_1, \ldots, X_n$ are a random sample from a $N(\mu, \sigma^2)$ distribution, μ is estimated with the sample mean $\overline{X}$ and σ^2 is estimated with the sample variance s^2

$$\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \qquad s^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i - \overline{X} \right)^2$$

unbiased estimator

asymptotically unbiased estimator ?

- As with any sample mean, the uncertainty in $\bar{X}$ is $\sigma / \sqrt{n}$ which we replace with $s / \sqrt{n}$, if σ is unknown. The mean is an unbiased estimator of μ.

Sample Variance is an Asymptotically Unbiased Estimator (1/1)

- Sample variance s^2 is an asymptotically unbiased estimator of the population variance σ^2

$$
\begin{aligned}
E\left[s^2\right] &= E\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2\right] \\
&= E\left[\frac{1}{n}\sum_{i=1}^{n}\left(X_i^2 - 2X_i\cdot\overline{X} + \overline{X}^2\right)\right] \\
&= E\left[\frac{\left(\sum_{i=1}^{n} X_i^2\right) - 2n\cdot\overline{X}^2 + n\overline{X}^2}{n}\right] \\
&= E\left[\frac{\left(\sum_{i=1}^{n} X_i^2\right) - n\cdot\overline{X}^2}{n}\right] \\
&= \frac{\left(\sum_{i=1}^{n} E\left[X_i^2\right]\right) - n\cdot E\left[\overline{X}^2\right]}{n}
\end{aligned}
$$

$$s^2 = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

$$\sum_{i=1}^{n} X_i = n\cdot\overline{X}$$

Sample Variance is an Asymptotically Unbiased Estimator (2/2)

$$\operatorname{Var}(\bar{X}) = \frac{\sigma^2}{n} = E[\bar{X}^2] - (E[\bar{X}])^2$$
$$\Rightarrow E[\bar{X}^2] = \frac{\sigma^2}{n} + (E[\bar{X}])^2 = \frac{\sigma^2}{n} + \mu^2$$

$$\mathbf{E}[s^2] = \frac{\left(\sum_{i=1}^{n} \mathbf{E}[X_i^2]\right) - n \cdot \mathbf{E}[\bar{X}^2]}{n}$$
$$= \frac{n(\sigma^2 + \mu^2) - n\left(\frac{\sigma^2}{n} + \mu^2\right)}{n}$$
$$= \frac{(n-1)}{n}\sigma^2 \xrightarrow{n=\infty} \sigma^2$$

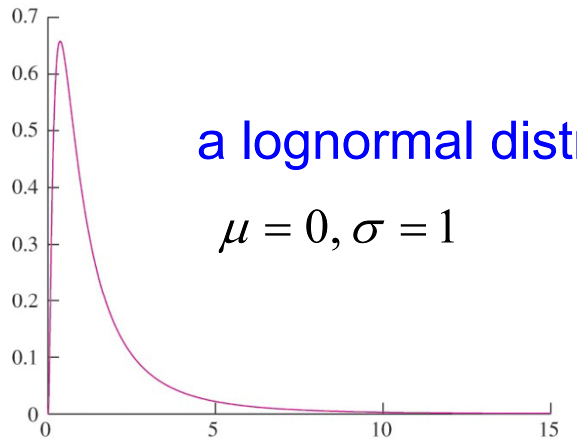
$$\operatorname{Var}(X_i) = \sigma^2 = \mathbf{E}[X_i^2] - (\mathbf{E}[X_i])^2$$
$$\Rightarrow \mathbf{E}[X_i^2] = \sigma^2 + (\mathbf{E}[X_i])^2 = \sigma^2 + \mu^2$$

The size of the observed sample

The Lognormal Distribution

- For data that contain outliers (on the right of the axis), the normal distribution is generally not appropriate. The lognormal distribution, which is related to the normal distribution, is often a good choice for these data sets
- If $X \sim N(\mu, \sigma^2)$, then the random variable $Y = e^X$ has the lognormal distribution with parameters μ and σ^2
- If Y has the lognormal distribution with parameters μ and σ^2, then the random variable $X = \ln Y$ has the $N(\mu, \sigma^2)$ distribution

Probability Density Function



a lognormal distribution with parameters

$\mu = 0, \sigma = 1$

pdf, Mean and Variance of Lognormal

- The pdf of a lognormal random variable with parameters μ and σ^2 is

$$f(y) = \begin{cases} \frac{1}{\sigma y \sqrt{2\pi}} \exp\left[-\frac{1}{2\sigma^2} (\ln y - \mu)^2 \right], & y > 0 \\ 0, & \text{Otherwise} \end{cases}$$

- The mean $E(Y)$ and variance $V(Y)$ are given by

$$E(Y) = e^{\mu + \sigma^2/2} \qquad V(Y) = e^{2\mu + 2\sigma^2} - e^{2\mu + \sigma^2}$$

  – Can be shown by advanced methods

pdf, Mean and Variance of Lognormal

- Recall "Derived Distributions"

$$Y = e^X,\ X \sim N\left(\mu, \sigma^2\right) \Rightarrow f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$$F_Y(y) = P(Y \le y) = P\left(e^X \le y\right) = P(x \le \log y) = F_X(\log y)$$

$$\Rightarrow$$

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \frac{dF_X(\log y)}{d \log y} \frac{\log y}{dy}$$

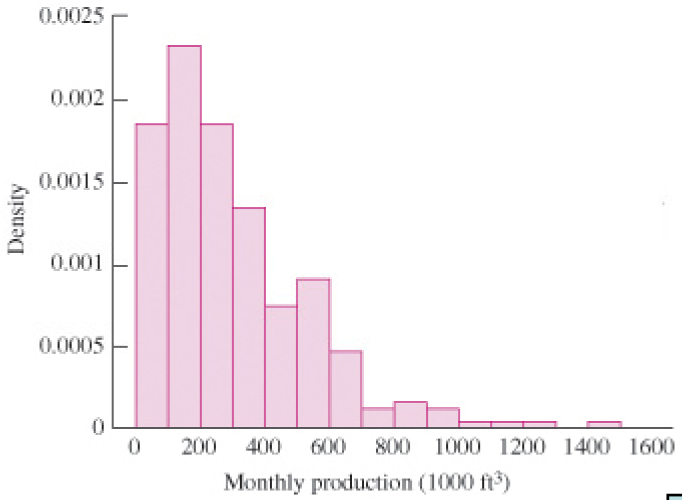
$$= f_X(\log y) \cdot \frac{1}{y}$$

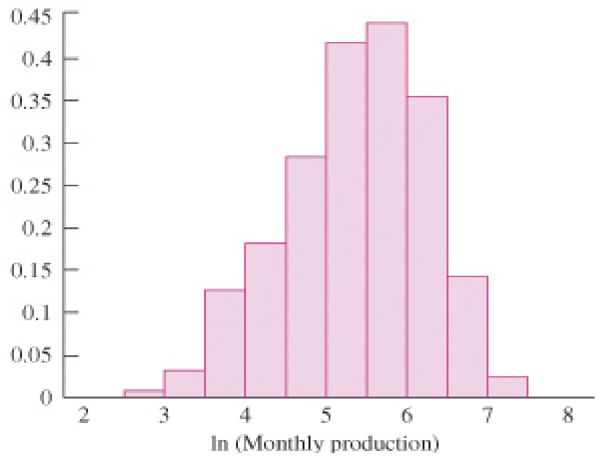
$$= \frac{1}{y\sqrt{2\pi}\,\sigma} \exp\left[-\frac{1}{2\sigma^2}(\log y - \mu)^2\right]$$

Test of “Lognormality”

- Transform the data by taking the natural logarithm (or any logarithm) of each value
- Plot the histogram of the transformed data to determine whether these logs come from a normal population

Probability Histogram





taking the natural logarithm on the values of the data

The Exponential Distribution

- The exponential distribution is a continuous distribution that is sometimes used to model the time that elapses before an event occurs
  - Such a time is often called a waiting time

- The probability density of the exponential distribution involves a parameter, which is a positive constant λ whose value determines the density function's location and shape

- We write $X \sim \text{Exp}(\lambda)$

pdf, cdf, Mean and Variance of Exponential

- The pdf of an exponential r.v. is

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, x > 0 \\ 0, \quad \text{otherwise} \end{cases}.$$

- The cdf of an exponential r.v. is

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

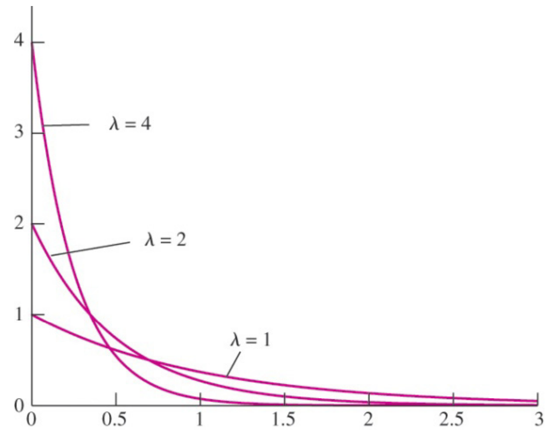
- The mean of an exponential r.v. is

$$\mu_X = 1/\lambda.$$

- The variance of an exponential r.v. is

$$\sigma_X^2 = 1/\lambda^2.$$

Probability Density Function



Lack of Memory Property for Exponential

- The exponential distribution has a property known as the lack of memory property: If $T \sim \text{Exp}(\lambda)$, and t and s are positive numbers, then

$$P(T > t + s \mid T > s) = P(T > t)$$

$$\begin{aligned}
P(T > t + s \mid T > s) &= \frac{P((T > t + s) \cap (T > s))}{P(T > s)} \\
&= \frac{P(T > t + s)}{P(T > s)} = \frac{1 - F_T(t + s)}{1 - F_T(s)} \\
&= \frac{e^{-\lambda(t+s)}}{e^{-\lambda s}} = e^{-\lambda t} = 1 - F_T(t) \\
&= P(T > t)
\end{aligned}$$

Estimating the Parameter of Exponential

- If $X_1,\ldots,X_n$ are a random sample from Exp(λ), then the parameter λ is estimated with $\hat{\lambda} = 1/\overline{X}$. This estimator is biased. This bias is approximately equal to λ/n (specifically, $\mu_{\hat{\lambda}} \approx \lambda + \lambda/n$). The uncertainty in $\hat{\lambda}$ is estimated with

$$\sigma_{\hat{\lambda}} = \frac{1}{\overline{X}\sqrt{n}}.$$

$$\sigma_{\hat{\lambda}} \approx \left| \frac{d}{d\overline{X}} \left(\frac{1}{\overline{X}} \right) \right| \sigma_{\overline{X}} = \frac{1}{\overline{X}^2} \cdot \sigma_{\overline{X}}$$

and $\sigma_{\overline{X}} = \frac{\sigma_X}{\sqrt{n}} = \frac{1}{\lambda} \cdot \frac{1}{\sqrt{n}} \approx \frac{\overline{X}}{\sqrt{n}}$

- This uncertainty estimate is reasonably good when the sample size n is more than 20

The Gamma Distribution (1/2)

- Let's consider the gamma function
  - For $r > 0$, the gamma function is defined by

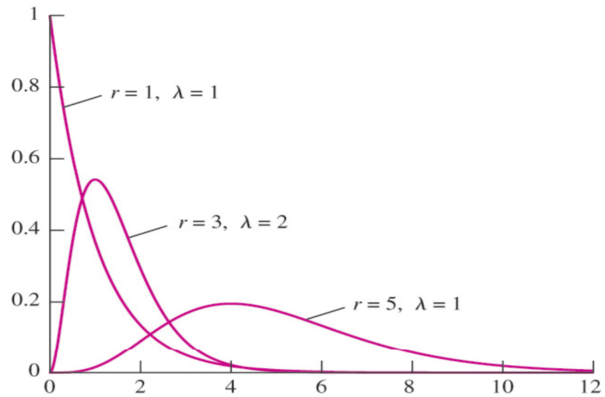
$$\Gamma(r) = \int_0^{\infty} t^{r-1} e^{-t} dt$$

  - The gamma function has the following properties:
    - If r is any integer, then $\Gamma(r) = (r-1)!$
    - For any r, $\Gamma(r+1) = r\,\Gamma(r)$
    - $\Gamma(1/2) = \sqrt{\pi}$

The Gamma Distribution (2/2)

- The pdf of the gamma distribution with parameters $r > 0$ and $\lambda > 0$ is

$$f(x) = \begin{cases} \dfrac{\lambda x^{r-1} e^{-\lambda x}}{\Gamma(r)}, x > 0 \\ 0, \quad x \le 0 \end{cases}.$$



- The mean and variance of Gamma distribution are given by
  - $\mu_X = r/\lambda$ and $\sigma_X^2 = r/\lambda^2$, respectively

- If $X_1, \ldots, X_r$ are independent random variables, each distributed as Exp(λ), then the sum $X_1 + \ldots + X_r$ is distributed as a gamma random variable with parameters r and λ, denoted as $\Gamma(r, \lambda)$

The Weibull Distribution (1/2)

- The Weibull distribution is a continuous random variable that is used in a variety of situations
- A common application of the Weibull distribution is to model the lifetimes of components
- The Weibull probability density function has two parameters, both positive constants, that determine the location and shape. We denote these parameters α and β
- If $\alpha = 1$, the Weibull distribution is the same as the exponential distribution with parameter $\lambda = \beta$

The Weibull Distribution (2/2)

- The pdf of the Weibull distribution is

$$f(x)=\begin{cases}\alpha\beta^{\alpha}x^{\alpha-1}e^{-(\beta x)^{\alpha}}, x>0\\ 0, \quad x\le 0\end{cases}.$$

- The mean of the Weibull is

$$\mu_X=\frac{1}{\beta}\Gamma\left(1+\frac{1}{\alpha}\right).$$

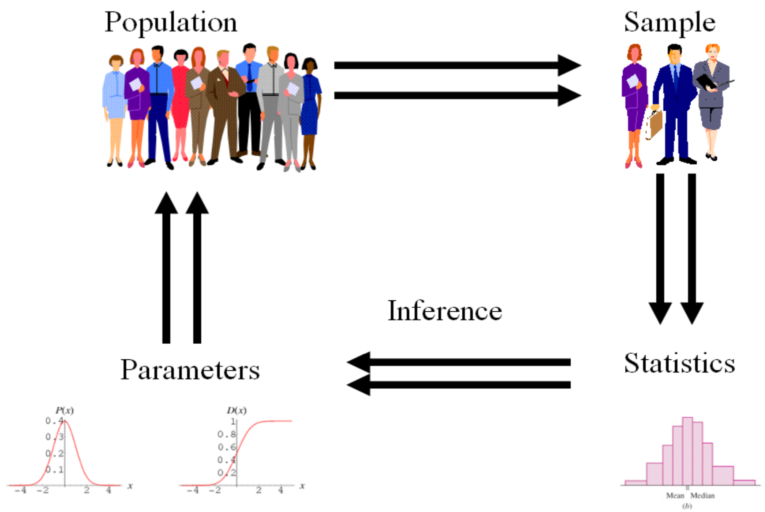
- The variance of the Weibull is

$$\sigma_X^2=\frac{1}{\beta^2}\left\{\Gamma\left(1+\frac{2}{\alpha}\right)-\left[\Gamma\left(1+\frac{1}{\alpha}\right)\right]^2\right\}.$$

Probability (Quantile-Quantile) Plots for Finding a Distribution

- Scientists and engineers often work with data that can be thought of as a random sample from some population
  - In many cases, it is important to determine the probability distribution that approximately describes the population

- More often than not, the only way to determine an appropriate distribution is to examine the sample to find a sample distribution that fits



Finding a Distribution (1/4)

- Probability plots are a good way to determine an appropriate distribution
- Here is the idea: Suppose we have a random sample $X_1, \ldots, X_n$
  - We first arrange the data in ascending order
  - Then assign increasing, evenly spaced values between 0 and 1 to each X_i
    - There are several acceptable ways to this; the simplest is to assign the value $(i - 0.5)/n$ to X_i order statistics

- The distribution that we are comparing the X's to should have a mean and variance that match the sample mean and variance
  - We want to plot $(X_i, F(X_i))$, if this plot resembles the cdf of the distribution that we are interested in, then we conclude that that is the distribution the data came from

Finding a Distribution (2/4)

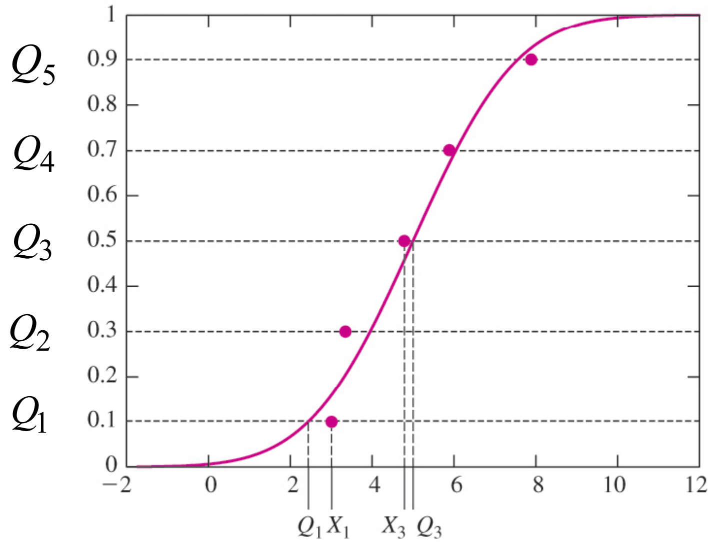
- Example: Given a sample X_i's arranged in increasing order

3.01, 3.35, 4.79, 5.96, 7.89

i	X_i	$(i-0.5)/5$
1	3.01	0.1
2	3.35	0.3
3	4.79	0.5
4	5.96	0.7
5	7.89	0.9

sample mean $\overline{X} = 5.00$

sample standard deviation $s = 2.00$

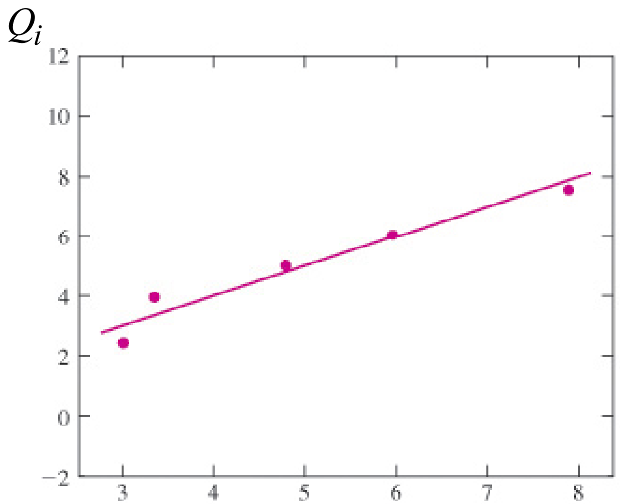


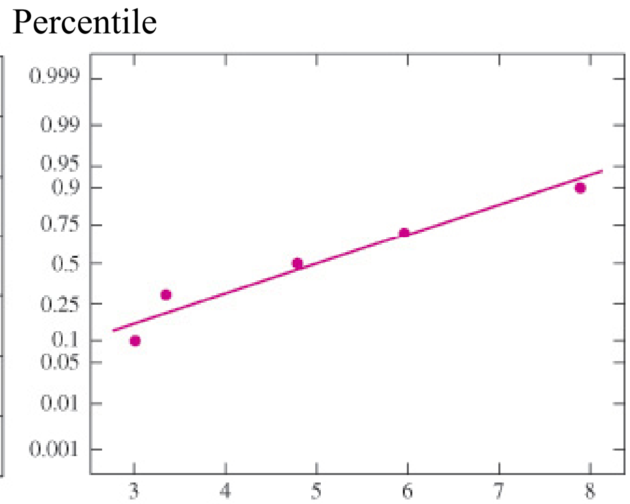
The curve is the cdf of $N(5,2^2)$. If the sample points Xi's came from the distribution, they are likely to lie close to the curve.

Finding a Distribution (3/4)

- When you use a software package, then it takes the (i – 0.5)/n assigned to each X_i and calculates the quantile (Q_i) corresponding to that number from the distribution of interest. Then it plots each (X_i, Q_i), or (Empirical Quantile, Quantile)
  - E.g., for the previous example (normal probability plot)

i	X_i	Q_i
1	3.01	2.44
2	3.35	3.95
3	4.79	5.00
4	5.96	6.05
5	7.89	7.56

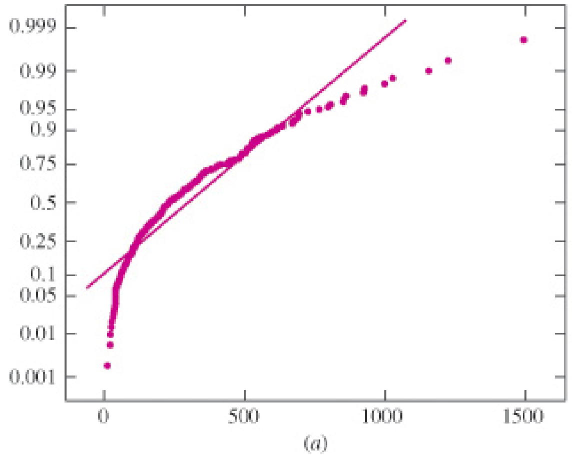




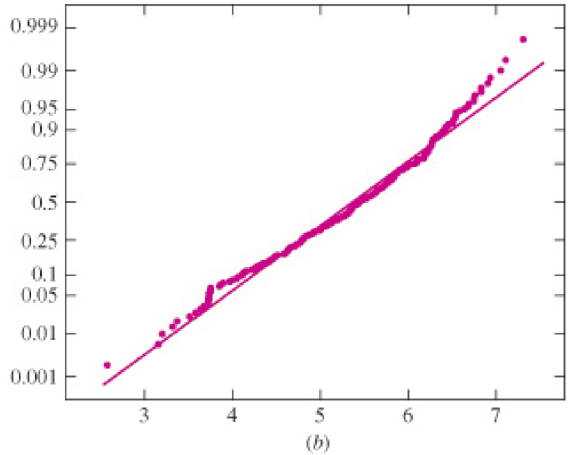
  - If this plot is a reasonably straight line then you may conclude that the sample came from the distribution that we used to find quantiles

Finding a Distribution (4/4)

- A good rule of thumb is to require at least 30 points before relying on a probability plot
  - E.g., a plot of the monthly productions of 255 gas wells



monthly productions



natural logs of monthly productions

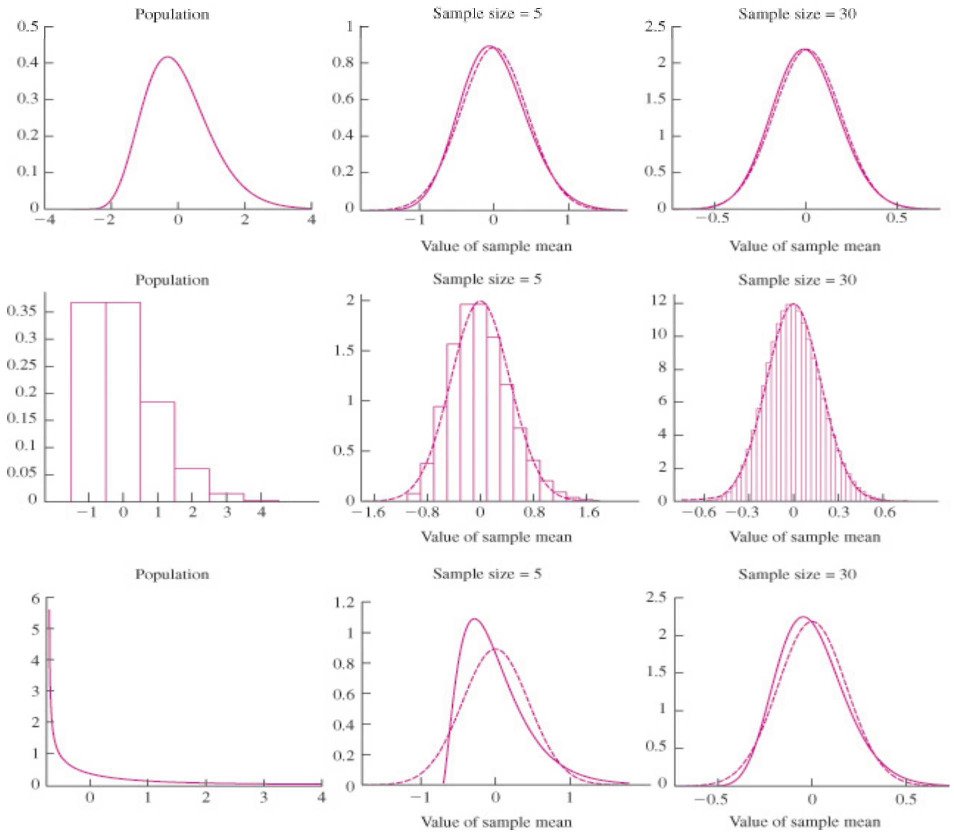
- The monthly productions follow a lognormal distribution !

The Central Limit Theorem (1/3)

- The Central Limit Theorem
  - Let $X_1,\ldots,X_n$ be a random sample from a population with mean μ and variance σ^2 (n is large enough)
  - Let $\overline{X} = \dfrac{X_1 + \cdots + X_n}{n}$ be the sample mean
  - Let $S_n = X_1+\ldots+X_n$ be the sum of the sample observations. Then if n is sufficiently large,
    - $\overline{X} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$ sample mean is approximately normal !
    - And $S_n \sim N(n\mu, n\sigma^2)$ approximately

The Central Limit Theorem (2/3)

- Example



continuous — Solid lines: real distributions of sample mean
Dashed lines: normal distributions

discrete — somewhat skewed to the right

continuous — highly skewed to the right

- Rule of Thumb
  – For most populations, if the sample size is greater than 30, the Central Limit Theorem approximation is good

The Central Limit Theorem (3/3)

- Example 4.64: Let X denotes the flaws in an 1in. length of copper wire, and its corresponding pmf, mean and variance are

x	$P(X = x)$
0	0.48
1	0.39
2	0.12
3	0.01

$$\mu = 0.66$$

$$\sigma^2 = 0.5244$$

  – One hundred wires are sampled from this population. What is the probability that the average number of flow per wire in this sample is less than 0.5 ?

$\Rightarrow$ Following the central limit theorem, we know that

the sample mean $\overline{X} \sim N(0.66, 0.005244)$

The z-score of $\overline{X} = 0.5$ is

$$z = \frac{0.5 - 0.66}{\sqrt{0.005244}} = -2.21 \qquad \therefore P(\overline{X} < 0.5) = P(Z < -2.21) = 0.0136$$

Law of Large Numbers

- Let $X_1, \ldots, X_n$ be a sequence of independent random variables with $\mathbf{E}[X_i] = \mu$ and $\mathrm{var}(X_i) = \sigma^2$.
  Let $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$. Then, for any $\varepsilon > 0$,

$$P\left(\left(\overline{X} - \mu\right) \geq \varepsilon\right) \leq \frac{\mathrm{var}\left(\overline{X}\right)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \quad \to 0, \quad \text{as } n \to 0$$

$$\mathbf{E}\left[\overline{X}\right] = \mathbf{E}\left[\frac{1}{n}\sum_{i=1}^{n} X_i\right] = \mu$$

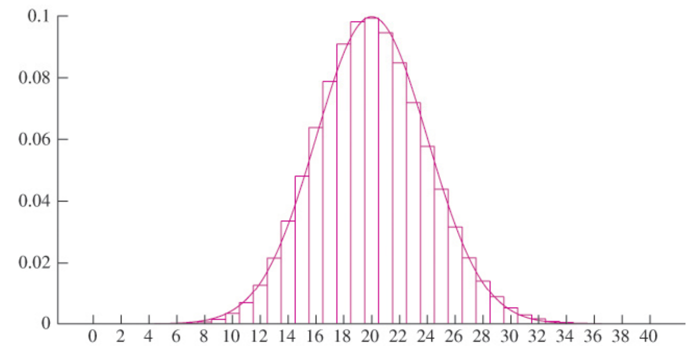
$$\mathrm{var}\left(\overline{X}\right) = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{var}(X_i) = \frac{\sigma^2}{n} \quad (\text{since } X_i \text{ are independent})$$

The desired result follows immediately from Chebyshev's inequality, which states that,

$$P\left(\left(X - \mu_X\right) \geq \varepsilon\right) \leq \frac{\sigma_X^2}{\varepsilon^2} \quad \text{for } \varepsilon > 0$$

Normal Approximation to the Binomial

- If $X \sim \text{Bin}(n,p)$ and if $np > 10$, and $n(1-p) > 10$, then
  - $X \sim N(np, np(1-p))$ approximately
  - And $\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$ approximately



Bin(100, 0.2) approximated by N(20, 16)

Recall that $X \sim \text{Bin}(n,p)$, then can X be represented as

$$X = Y_1 + Y_2 + \ldots + Y_n,$$

where $Y_1, Y_2, \ldots, Y$ is a sample from Bernoulli(p)

$\Rightarrow$ Following the central limit theorem, if n is large enough then

$$\hat{p} = \frac{X}{n} = \frac{Y_1 + Y_2 + \ldots + Y_n}{n} \text{ be approximated by } N\left(p, \frac{p(1-p)}{n}\right)$$

and

X can be approximated by $N(np, np(1-p))$

Normal Approximation to the Poisson

- Normal Approximation to the Poisson: If X ~ Poisson(λ), where λ > 10, then $X \sim N(\lambda, \lambda)$
  - The Poisson can be first approximated by Binomial and then by Normal

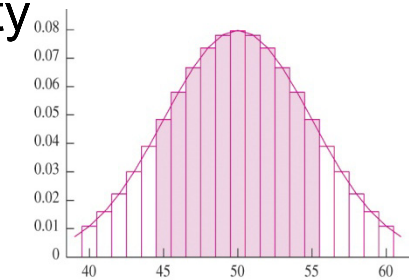
Note that variance of binomial :

$$\sigma^2 = np(1-p) = \lambda(1-p) \approx \lambda \quad (\text{if } p << 1)$$

Continuity Correction

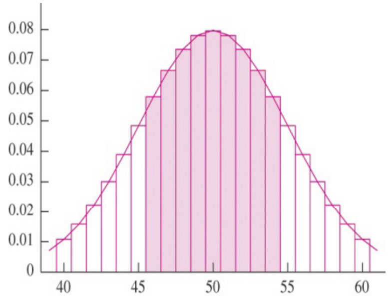
- The binomial distribution is discrete, while the normal distribution is continuous
- The continuity correction is an adjustment, made when approximating a discrete distribution with a continuous one, that can improve the accuracy of the approximation
  - If you want to include the endpoints in your probability calculation, then extend each endpoint by 0.5. Then proceed with the calculation

e.g., $P(45 \le X \le 55)$



  - If you want exclude the endpoints in your probability calculation, then include 0.5 less from each endpoint in the calculation

e.g., $P(45 < X < 55)$



Summary

- We considered various discrete distributions: Bernoulli, Binomial, Poisson, Hypergeometric, Geometric, Negative Binomial, and Multinomial
- Then we looked at some continuous distributions: Normal, Exponential, Gamma, and Weibull
- We learned about the Central Limit Theorem
- We discussed Normal approximations to the Binomial and Poisson distributions