Fundamentals of Business Statistics – PT IMBA 2007/8

PART B

Language of Probability

& Random Variables

Chapters 5 & 6

Course Outline

Graphs and tables	Chapter 2	Part A
Summary measures	Chapter 3	
Probability and probability distributions	Chapter 5	Part B
Normal, binomial, Poisson, and exponential distributions	Chapter 6	
Sampling & statistical inference	Chapters 8-11	Part C

Part A: Describe (Summarize) Data in Samples
Part B: Language of Probability & Random Variables
Part C: Sampling & Statistical Inference

"We may at once admit that any inference from the particular to the general must be attended with some degree of uncertainty, but this is not the same as to admit that such inference cannot be absolutely rigorous, for the nature and degree of the uncertainty may itself be capable of rigorous expression,"
R.A. Fisher in *The Design of Experiments*

Overview of Part B – Specifics

- **Probability density functions**
- The normal distribution family
- The binomial distribution family
- The Poisson distribution family
- The Exponential distribution family

Probability Density Functions

- A **probability density function**, usually denoted $f(x)$, *specifies* the probability distribution of a continuous random variable X.
- For a discrete random variable, the concepts of probability distribution (function) and probability density function are equivalent. This is not the case for a continuous random variable.
- A density value $f(x)$ assigned to a possible value x of a continuous random variable X merely specifies a *relative likelihood,* relative to the likelihood of other possible values. Though it resembles a histogram, **density values are not probabilities.**

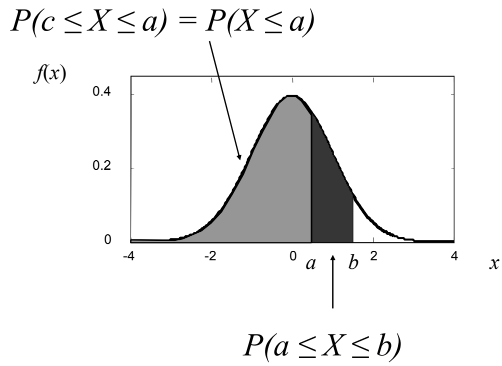
Probability Density Functions (cont.)

- In fact, the probability of any value x of a truly continuous random variable X, i.e. X can take on an indefinite number of different values, is **0**!
  (cf. law of large numbers)
- Hence, for a continuous random variable it only makes sense to **specify probabilities for value ranges** e.g. $P(a \leq X \leq b)$ for a variable X with *universe of discourse* $[c,d]$ *and* $c \leq a \leq b \leq d$. Note: c and d may be $-\infty$ and $+\infty$, respectively.

Probability Density Functions (cont.)

- The **area under the density function** bounded by a to the left and b to the right is, *per definition of* $f(x)$, equal to $P(a \leq X \leq b)$.
- Any function $f(x)$ can (*technically*) be a density function for a variable X with universe of discourse $[c,d]$ as long as it satisfies the following conditions:
  - $f(x) \geq 0$ for all $x \in [c,d]$.
  - The total area under the density function equals 1, i.e. the total probability of 1 is 'spread' over the value continuum under the curve.

Probability Density Functions (cont.)



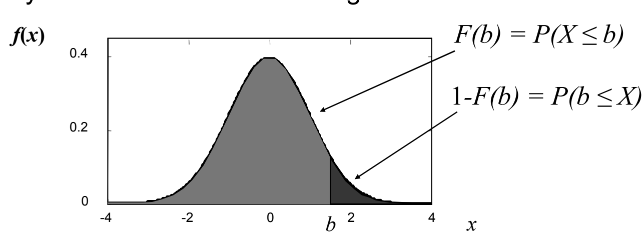
Exercise: Specify the density function for *Rand()* in Excel. (Hint: Uniform distribution over [0,1].)

Probability Density Functions (cont.)

- **The cumulative distribution function** of a continuous random variable X with universe of discourse $[c,d]$, usually denoted $F(x)$, gives the **probability** of observing a value of X less than or equal to x.

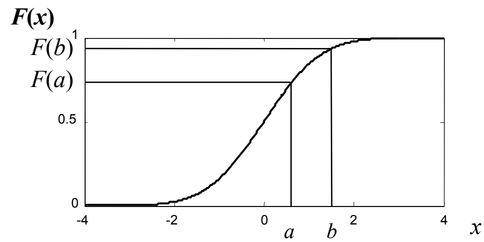
$$F(x) = P(c \leq X \leq x) = P(X \leq x)$$

- This probability equals the area under the density curve bounded by c to the left and x to the right.



Probability Density Functions (cont.)

$F(x) = P(X \leq x)$



$P(a \leq X \leq b)=F(b)-F(a)$

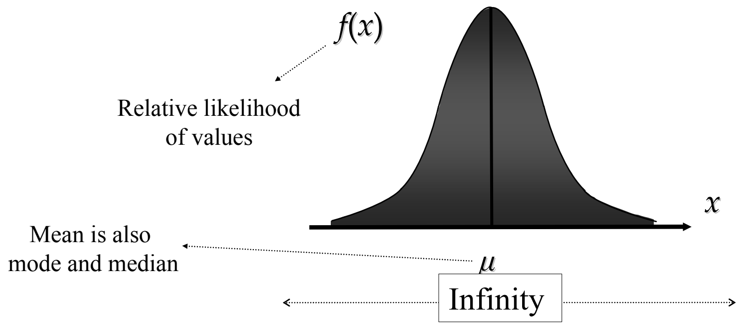
Exercise: Specify the cumulative distribution function for *Rand()* in Excel.

Overview of Part B – Specifics

- Probability density functions
- **The normal distribution family**
- The binomial distribution family
- The Poisson distribution family
- The Exponential distribution family

Normal Distribution

- The **single most important** distribution in statistics is the **normal** distribution. It arises naturally in many social, economical, physical, and biological measurement situations. It is also the cornerstone of statistical inference.
- The normal distribution is a *continuous* distribution and is associated with the familiar *symmetric bell-shaped* density curve.



Normal Distribution (cont.)

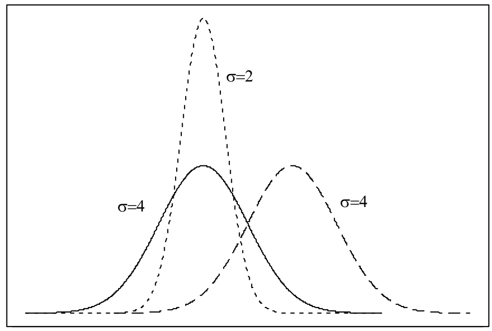
- A continuous, symmetric, bell-shaped curve that complies with the properties of a density function for a continuous random variable X with possible value range $[-\infty, +\infty]$ is completely described by the following formula:

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad \text{for } -\infty \le x \le +\infty$$

- Actually, this describes a complete class (or family) of distributions; one for each choice of the mean μ and standard deviation σ. (Note: e = 2.71828…)
- A shorthand notation for stating that a variable X is normally distributed with mean μ and standard deviation σ *is* $X \sim N(\mu,\sigma)$.

Normal Distribution (cont.)

- Shape:
  – Changing the mean (ceteris paribus) shifts the curve to the left or right.
  – Changing the standard deviation (ceteris paribus) widens or tightens the curve.

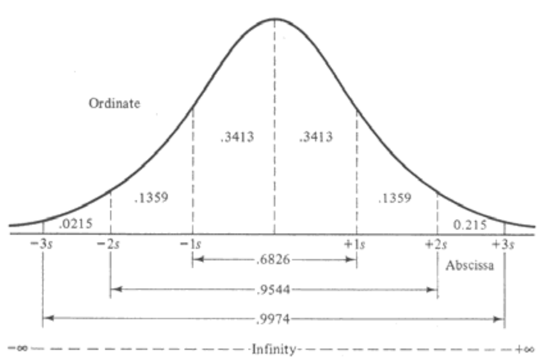


Normal Distribution (cont.)

- Characteristic to any of these shapes:

$$P(\mu - 1\sigma \le X \le \mu + 1\sigma) = 0.6826$$
$$P(\mu - 2\sigma \le X \le \mu + 2\sigma) = 0.9544$$
$$P(\mu - 3\sigma \le X \le \mu + 3\sigma) = 0.9974$$

Remember the ‘empirical rules’ in Part A?



Standard Normal Distribution

- There are infinitely many normal distributions. However we single out one of these for special attention, the **standard normal** distribution.
- This distribution has a mean 0 and standard deviation 1, so we can denote it by $N(0,1)$.
- It is also known as the Z-distribution.
- **Standardizing** a variable involves subtracting its mean and then dividing the difference by the standard deviation.

Standard Normal Distribution (cont.)

Suppose $X \sim N(\mu,\sigma)$ and $Z = \dfrac{X-\mu}{\sigma}$
then $Z \sim N(0,1)$.

(1) Since X has value range $[-\infty,+\infty]$, Z has this range too.
(2) From Part B - Generics : $E(Z) = 0$ and $\sigma(Z) = 1$.
(3) If Z were normally distributed with $E(Z) = 0$ and $\sigma(Z) = 1$ then
f(z) would have functional form :

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

This can easily be verified by substituting for z in *f(x)* and making use of the result in (2)

.

Note : only standardizing a normally distributed variable results in a standard normal variable!

Standard Normal Distribution (cont.)

- One reason for **standardizing** is to measure variables with different means and/or standard deviations on a single scale.
- It is also easy to **interpret** a Z-value. It is the number of standard deviations to the right or left of the mean.
- If Z is positive, the original score is to the right of the mean, if Z is negative, the original score is to the left. Note: centered at 0.

$$P(\mu - 1\sigma \le X \le \mu + 1\sigma) = 0.6826$$
$$P(\mu - 2\sigma \le X \le \mu + 2\sigma) = 0.9544$$
$$P(\mu - 3\sigma \le X \le \mu + 3\sigma) = 0.9974$$

↓

$$P(-1 \le \frac{X-\mu}{\sigma} \le 1) = 0.6826$$
$$P(-2 \le \frac{X-\mu}{\sigma} \le 2) = 0.9544$$
$$P(-3 \le \frac{X-\mu}{\sigma} \le 3) = 0.9974$$

↓

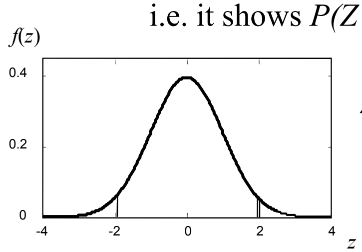
$$P(-1 \le Z \le 1) = 0.6826$$
$$P(-2 \le Z \le 2) = 0.9544$$
$$P(-3 \le Z \le 3) = 0.9974$$

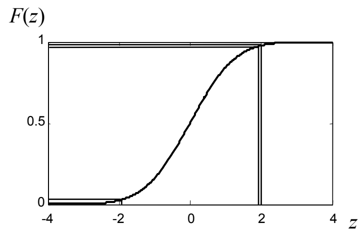
Standard Normal Distribution (cont.)

- A common use for Z-values and the standard normal distribution is in **calculating probabilities** and **percentiles** by the 'traditional' method, which is based on a table of the standard normal distribution found in most textbooks.
- The body of the table contains the probabilities for $P(Z \le z)$. The left and top margins contain possible values of Z. For looking up these probabilities we read the table '*from margins to body*'.
- If we are given a probability, we can use the table to find the value with this much probability to the left of it under the standard normal curve. This is called a **percentile** calculation. For looking up these values we read the table '*from body to margins*'.
- Sometimes 'linear interpolation' is needed. Excel does this automatically.

Standard Normal Distribution (cont.)

Using the cumulative probability table for Z, i.e. it shows $P(Z \le z)$:





?

(1) What is $P(Z \le 2)$*?*

(2) What is the Z range that is about 95% most likely?

z	.00	.01	.02	.03	.04	.05	.06
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948

Normal Calculations

- Basically, there are two types of calculations we typically make with any normal distribution, i.e.
  - finding probabilities
  - finding percentiles
- Excel makes each of these simple.

Finding Probabilities

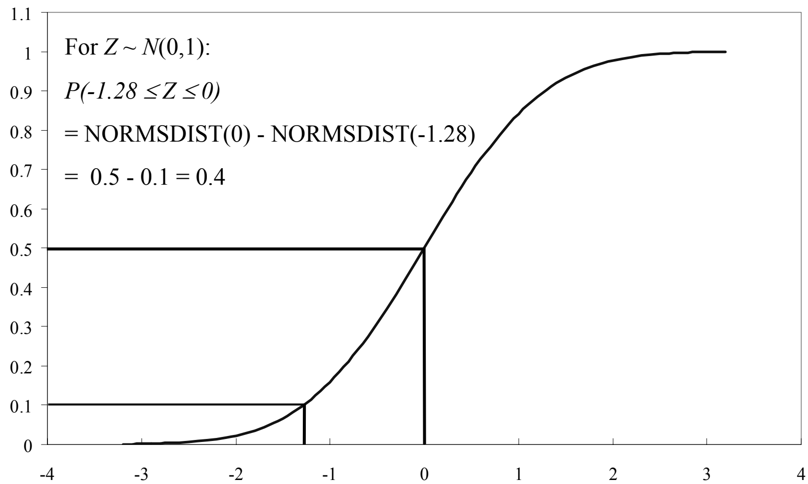
- To find probabilities using a normally distributed variable X we would calculate areas under its probability density function $f(x)$. (~calculus!)
- It is easier to use its cumulative distribution function $F(x)=P(X \leq x)$.
- Excel has two functions for $F(x)$.
  - If $X \sim N(\mu,\sigma)$ then use $F(x) =$ NORMDIST(x,μ,σ,1).
  - If $X \sim N(0,1)$ then use $F(x) =$ NORMSDIST(x).

Finding Probabilities (cont.)

- If $X \sim N(\mu,\sigma)$ then
  - $P(X \leq a)$ = NORMDIST(a,μ,σ,1).
  - $P(X > a)$ = 1 - NORMDIST(a,μ,σ,1).
  - $P(a \leq X \leq b)$ = NORMDIST(b,μ,σ,1) - NORMDIST(a,μ,σ,1).
- If $X \sim N(0,1)$ then
  - $P(X \leq a)$ = NORMSDIST(a).
  - $P(X > a)$ = 1 - NORMSDIST(a).
  - $P(a \leq X \leq b)$ = NORMSDIST(b) - NORMSDIST(a).

Note: if we can re-express as $P(X \leq a)$ we can calculate it!

Finding Probabilities Example 1



Finding Probabilities Example 2

- Example 6.3 p. 250
- Company "ZTel" hires based on test scores that are assumed to be normally distributed. *Note: we are using a histogram to guide us to this assumption (cf. law of large numbers).*

	A	B	C	D	E	F
1	**Personnel Accept/Reject Example**					
2						
3	Mean of test scores	525		**Range names**		
4	Stdev of test scores	55		Mean: B3		
5				Stdev: B4		
6	**Current Policy**					
7	Automatic accept point	600				
8	Automatic reject point	425				
9						
10	Percent accepted	8.63%		1-NORMDIST(B7,Mean,Stdev,1)		
11	Percent rejected	3.45%		NORMDIST(B8,Mean,Stdev,1)		
12						

What are resp. the accept and reject probabilities if they set their policy this way?

Answer!

Finding Percentiles

- Sometimes we want to find the value of a normally distributed variable X with a certain probability to the left of it. These are the percentile values.
- Excel has two functions for percentile values x.
  - If $X \sim N(\mu,\sigma)$ then use x = NORMINV($P(X \le x)$,μ,σ).
  - If $X \sim N(0,1)$ then use x = NORMSINV($P(X \le x)$).

Finding Percentiles Example 1

- Example 6.3 p. 250 (continued)
- Company "ZTel" hires based on test scores that are assumed to be normally distributed normally distributed. Based on the law of large numbers they modeled this distribution.

	A	B	C	D	E	F
1	**Personnel Accept/Reject Example**					
2						
3	Mean of test scores	525		**Range names**		
4	Stdev of test scores	55		Mean: B3		
5				Stdev: B4		
6	**Current Policy**					
7	Automatic accept point	600				
8	Automatic reject point	425				
9						
10	Percent accepted	8.63%		1-NORMDIST(B7,Mean,Stdev,1)		
11	Percent rejected	3.45%		NORMDIST(B8,Mean,Stdev,1)		
12						
13	**New Policy**					
14	Percent accepted	15%				
15	Percent rejected	10%				
16						
17	Automatic accept point	582		NORMINV(1-B14,Mean,Stdev)		
18	Automatic reject point	455		NORMINV(B15,Mean,Stdev)		

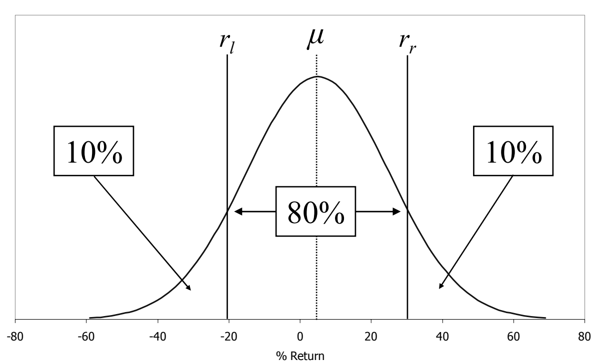
What would be the accept and reject test scores if these new probabilities were to define their new policy?

Answer!

Finding Percentiles Example 2

- Suppose the distribution of the annual return on Microsoft's stock is $R \sim N(5,20)$. So, let μ = 5 and σ = 20.
- What values of the return would you expect to observe less than 20% of the time?

Finding Percentiles Example 2 (cont.)



(1) We know that the tails are the most unlikely.
(2) Due to symmetry, outer left 10% equally unlikely as outer right 10%. Thus, identify r_l and r_r.

Finding Percentiles Example 2 (cont.)

- Hence, we seek values r_l and r_r such that $P(r_l \leq R \leq r_r) = 0.8$.
- To find r_l we find $P(R \leq r_l) = 0.1$ (i.e. lower 10% of left tail values)
  - NORMINV(0.1, 5, 20) gives $r_l = -20.631$.
- To find r_r we find $P(R \leq r_r) = 0.9$ (i.e. upper 10% of right tail values)
  - NORMINV(0.9, 5, 20) gives $r_r = 30.631$.
- So, the values of R in the range [-20.631, 30.631] cover the 80% most likely value occurrences; the values beyond the boundaries cover the 20% least likely value occurrences.

Finding Percentiles Example 2 (cont.)

- We can also use the standard normal or Z-distribution to solve this problem.
- We are still after values r_l and r_r such that $P(r_l \leq R \leq r_r) = 0.8$.
- 'Standardizing' this normal variable problem gives:
  - $P(r_l \leq R \leq r_r) = P((r_l - \mu) / \sigma \leq Z \leq (r_r - \mu) / \sigma) = 0.8$ or
  - $P(r_l \leq R \leq r_r) = P(z_l \leq Z \leq z_r) = 0.8$
- To find z_l a we compute NORMSINV(0.1) = -1.28.
- Because the mean of the Z-distribution is zero, $z_r = - z_l = 1.28$.

Finding Percentiles Example 2 (cont.)

- To find the values r_l and r_r we transform the *Z*-values back into R values.

$z_l = (r_l - \mu) / \sigma = -1.28$ **OR**
$r_l = \mu - 1.28\ \sigma = 5 - 1.28\ (20) = -20.631$

$z_r = (r_r + \mu) / \sigma = +1.28$ **OR**
$r_r = \mu + 1.28\ \sigma = 5 + 1.28\ (20) = 30.631$

Overview of Part B – Specifics

- Probability density functions
- The normal distribution family
- **The binomial distribution family**
- The Poisson distribution family
- The Exponential distribution family

All are related to special types of random variables.

Binomial Distribution

- The **binomial distribution** is a **discrete** distribution for the **number of successes** X in 2 situations (*For which experiment?*):
  - whenever we perform a **sequence** of probabilistically **independent, identical** sub-experiments, each of which has only two possible outcomes, i.e. success or failure. Or, alternatively, …
  - whenever we 'sample' (with replacement!) from a population with only two types of members (e.g. males and females)
- A single such sub-experiment is called a '**Bernoulli trial**':
  - Each sub-experiment results in either success or failure
  - Independent → outcome of a trial is independent from other trials
  - Identical → probability of success p stays the same across trials
- E.g. sequence of tossing coins, of values shown by a die

Binomial Distribution (cont.)

- Formally: The probability of observing x successes for n Bernoulli trials, where the probability of success on each trial is p, is:

$$p(x) = \frac{n!}{x!(n-x)!} p^x (1-p)^{(n-x)} \qquad \text{for} \quad x = 0,1,2,\ldots,n$$

- Note:
  - The second term of $p(x)$ is the probability of one particular sequential configuration of x successes and $(n\text{-}x)$ failures.
  - The first term of $p(x)$ is the number of sequential configurations (combinations) to have x successes and $(n\text{-}x)$ failures.

Binomial Distribution (cont.)

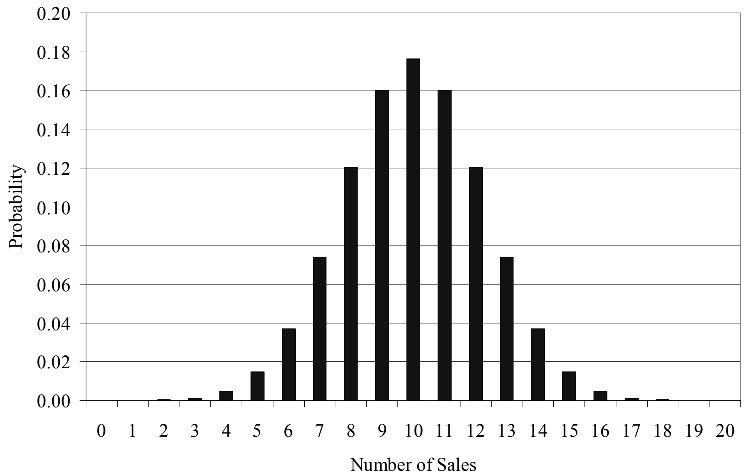
- Summary measures
  - Expected value: $E(X) = np$
  - Variance: $\sigma^2(X) = np(1-p)$
  - Standard deviation: $\sigma(X) = \sqrt{np(1-p)}$
- Excel
  - BINOMDIST(x,n,p,cum) where
    for $cum=0$, the function returns $p(x)$ ("exactly") and
    for $cum=1$, the function returns $P(X \le x)$ ("at most")
  - For percentiles we have CRITBINOM($n,p,P(X \le x)$) = x

Sales Call Example

- Suppose you will make 20 sales calls each day and that the probability of making a sale on any one call is 0.5.
- We ***assume*** the probability of a sale is independent of any previous sales call, and that this probability remains constant. *Is this justified???*
- The probability of making exactly 10 sales out of 20 sales calls is then

$$p(10) = \frac{20!}{10!(20-10)!} 0.5^{10} (1-0.5)^{(20-10)} = 0.1762$$

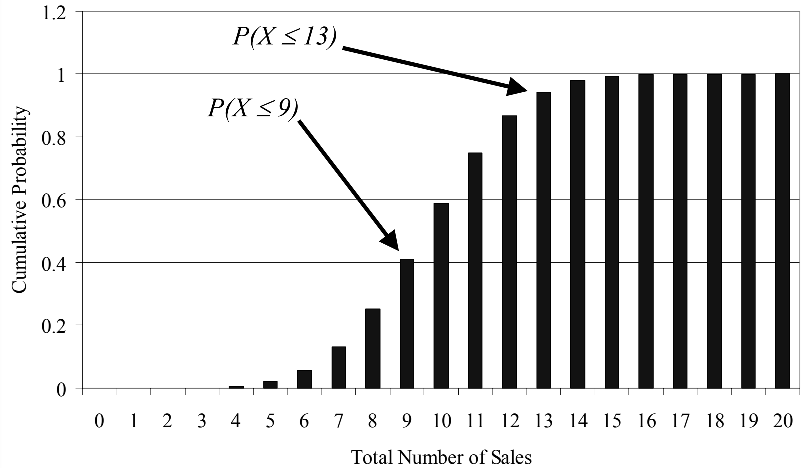
Sales Call Example (cont.)



Sales Call Example (cont.)

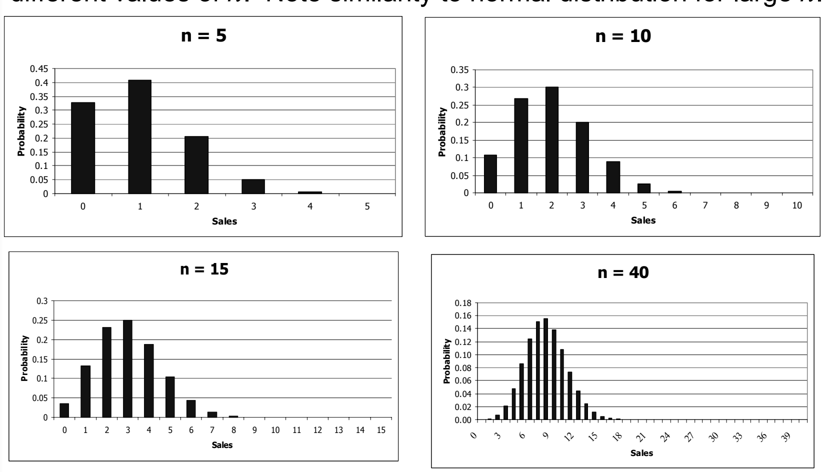
- What is probability of making between 10 and 13 sales?
- What is *P(X=10 or X=11 or X=12 or X=13)* or alternatively $P(10 \le X \le 13)$?
- Addition rule says: Sum the values BINOMDIST(x,20,0.5,0) for x=10,11,12,13.
- Use cumulative distribution function: BINOMDIST(13,20,0.5,1) – BINOMDIST(9,20,0.5,1)

Sales Call Example (cont.)



Family of Distributions

The binomial distribution depends on n and p. Consider $p=0.2$ for different values of n. Note similarity to normal distribution for large n.



Normal Approximation

- If n is large and p is not too close to 0 or 1, the distribution of the number of successes on n Bernoulli trials is <u>*approximated*</u> by the normal distribution.
- Experience indicates that the approximation is accurate as long as $np > 5$ for $p \leq 0.5$, and $n(1\text{-}p) > 5$ for $p > 0.5$.

Normal Approximation Example

- Let's use the normal approximation to calculate the probability of making <u>exactly</u> 10 sales on 20 calls.
- Check: $20(0.5) > 5$ so we satisfy the criterion.
- To use the normal distribution we MUST calculate an interval. **Why?**
  - It is common to add and subtract ½ to the value of interest. Hence $P(X=10)$ becomes $P(9.5 \leq X \leq 10.5)$.
  - We take $\mu=np=10$ and $\sigma^2=np(1\text{-}p)=5$
  - NORMDIST(10.5,10,2.24,1) – NORMDIST(9.5,10,2.24,1) = 0.1766 which compares with the binomial value of 0.1762.

Overview of Part B – Specifics

- Probability density functions
- The normal distribution family
- The binomial distribution family
- **The Poisson distribution family**
- The Exponential distribution family

Poisson Distribution

- The Poisson distribution is a discrete distribution for modeling the probability of a finite number of event type occurrences X in a *particular* interval of time or space:

$$p(x) = \frac{\lambda^x e^{-\lambda}}{x!} \quad \text{for } x = 0,1,2,...$$

- Summary measures:

This only makes sense if …

$$E(X) = \lambda \text{ and } \sigma^2(X) = \lambda$$

- Thus, parameter λ stands for the mean number of events during the particular time interval. (Note: e = 2.71828…)
- Excel: POISSON(x,λ,cum)

Poisson Distribution (cont.)

- E.g. number of telephone calls arriving per hour, number of defects per square inch of fabric.
- In general, time-based events are part of the theory of queuing, which encompasses modeling inventory issues.
- Implicit assumption: Whether the type of event of interest occurs in any particular interval is independent of whether it occurs in any other non-overlapping interval. (= ‘no memory’ feature)
- Approximation to the Binomial distribution:
  - If n is large and p is small, the probability of x successes in n Bernoulli trials can be approximated by a Poisson distribution with $\lambda = np$.
  - Experience indicates that the approximation is good if $n > 20$ and $p < 0.5$.

Example 6.13

Managing Inventory of Televisions at Kriegland

The Poisson Distribution

Background Information

- Krieglands is a department store that sells various brands of color television sets.
- One of the manager's biggest problems is to decide on an appropriate inventory policy for stocking television sets.
- On the one hand, he wants to have enough in stock so that customers receive their requests right away, but on the other hand, he does not want to tie up too much money in inventory that sits on the storeroom floor.

Background Information (cont.)

- Most of the difficulty results from the unpredictability of customer demand.
  - If this demand were constant and known, the manager could decide on an appropriate inventory policy fairly easily.
  - But the demand varies widely from month to month in a random manner.
- All the manager knows is that the historical average demand per month is approximately 17.

Background Information (cont.)

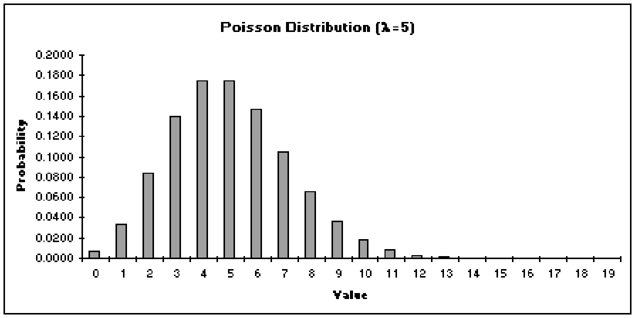
- Therefore, he decides to call in a consultant. The consultant immediately suggests using a probability model.
- Specifically, she attempts to find the probability distribution of demand in a typical month.
- How might she proceed?

  (Focus on the steps of the process towards a solutions!)

Solution

- Let X be the demand in a typical month.
- The consultant knows that there are many possible values of X.
- Historical records show that monthly demands have always been between 0 and 40; therefore almost all of the probability should be assigned to values between 0 and 40.
- She discovers from the manager that the histogram of demands from previous months is shaped much like the following graph.

Solution (cont.)



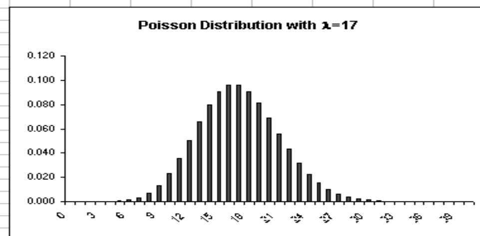
- The above histogram is a typical graph of the Poisson distribution.

Solution (cont.)

- Knowing that the Poisson distribution has the same basic shape, the consultant decides to model the monthly distribution with a Poisson distribution.
- To choose a particular Poisson distribution, all she has to do is choose a value of λ, the mean demand per month.
- Because the historical average is approximately 17, she chooses λ=17.
- Now she can test the Poisson model by calculating probabilities of various events, using POISSON(x, 17,1).

Solution (cont.)

	A	B	C	D
1	**Poisson distribution for department store example**			
2				
3	Mean monthly demand (λ)	17		**Range names:** MeanDem: B3
4				
5	Representative probability calculations			
6	Less than or equal to 20	0.805		=POISSON(20,MeanDem,1)
7	Between 10 and 15 (inclusive)	0.345		=POISSON(15,MeanDem,1)-POISSON(9,MeanDem,1)
8				
9	Individual probabilities			
10	Value	Prob		
11	0	0.000		=POISSON(A11,MeanDem,0)
12	1	0.000		
13	2	0.000		
14	3	0.000		
15	4	0.000		
16	5	0.000		
17	6	0.001		
18	7	0.003		
19	8	0.007		
20	9	0.014		
21	10	0.023		
22	11	0.036		
23	12	0.050		
24	13	0.066		
25	14	0.080		
26	15	0.091		
27	16	0.096		
28	17	0.096		
29	18	0.091		
30	19	0.081		
31	20	0.069		
32	21	0.056		
33	22	0.043		
34	23	0.032		
35	24	0.023		
36	25	0.015		
37	26	0.010		
38	27	0.006		
39	28	0.004		
40	29	0.002		
41	30	0.001		
42	31	0.001		
43	32	0.000		
44	33	0.000		
45	34	0.000		
46	35	0.000		
47	36	0.000		
48	37	0.000		
49	38	0.000		
50	39	0.000		
51	40	0.000		



Solution (cont.)

- Given these calculations she can ask the manager whether these probabilities are a reasonable approximation of reality.
- If the manager believes that these probabilities and other similar probabilities are reasonable, then the statistical part of the consultant’s job is finished.
- Otherwise, she must try a different Poisson distribution – a different value of λ – or perhaps a different type of distribution altogether.

Overview of Part B – Specifics

- Probability density functions
- The normal distribution family
- The binomial distribution family
- The Poisson distribution family
- **The Exponential distribution family**

Exponential Distribution

- If a process is characterized by a Poisson distribution then it can be shown that the **time in between consecutive event occurrences** X abides by an exponential distribution. The density function for this **continuous** variable X is:

$$f(x) = \lambda e^{-\lambda x} \quad \text{for } 0 \le x \le +\infty$$

where $\lambda > 0$ is the mean number of events per unit of time.

- Summary measures:

$$E(X) = \frac{1}{\lambda} \quad \text{and} \quad \sigma^2(X) = \frac{1}{\lambda}$$

- If λ is expressed as number of events per unit of time, how then is 1/ λ expressed?

Finding Probabilities

- As with any continuous variable X, we find probabilities by looking at areas under its density function, i.e. $P(a \leq X \leq b)$.
- It is often easier to work with the cumulative distribution function in these cases. For an exponentially distributed variable X this is

$$P(X \leq x) = (1 - e^{-\lambda x})$$

- Thus:

$$P(a \leq X \leq b) = (e^{-\lambda a} - e^{-\lambda b})$$

- In Excel $P(X \leq x)$ = EXPONDIST(x,λ,1)

Example

- Let the mean number of customers arriving in a shop be λ=6. So, the mean time in between arrivals is 1/6=0.1667 hours or one customer arrives every 10 minutes.
- What is the probability that the time interval between two consecutive arrivals is 10 minutes or less?

$$P(0 \leq X \leq 0.1667) = e^{-6(0)} - e^{-6(0.1667)}$$
$$P(0 \leq X \leq 0.1667) = 1 - e^{-1}$$
$$P(0 \leq X \leq 0.1667) = 0.632$$

- Between 5 and 15 minutes?

$$P(\frac{5}{60} \leq X \leq \frac{15}{60}) = e^{-6(\frac{5}{60})} - e^{-6(\frac{15}{60})}$$
$$P(\frac{5}{60} \leq X \leq \frac{15}{60}) = e^{-0.5} - e^{-1.5}$$
$$P(\frac{5}{60} \leq X \leq \frac{15}{60}) = 0.606 - 0223 = 0.383$$

Exponential Functions for λ=6