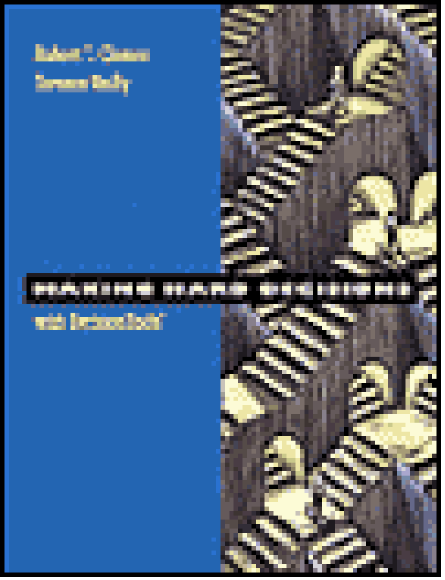


Making Hard Decisions

R. T. Clemen, T. Reilly

Chapter 9

Theoretical Probability Models

Draft: Version 1

Theoretical Models Applied

Theoretical Probability Models may be used when they describe the physical model **"adequately"**

Examples:

1. The outcome of an IQ test - **Normal Distribution**.
2. The lifetime of a component exhibiting aging - **Weibull Distribution**.
3. The length of a telephone call - **Exponential distribution.**
4. The time between two people arriving at a post office - **Exponential distribution**.
5. The number of people arriving at a post office in one hour - **Poisson Distribution**.
6. The number of defectives in releasing a batch of fixed size - **Binomial distribution**.

Draft: Version 1

The Binomial Distribution

Assumptions:

1. A fixed number of trials, say N.
2. Each trial results in a "Success" or "Failure"
3. Each Trial has the same probability of success p.
4. Different Trials are independent.

Define:

X = "# Successes in a sequence of N trials",

$$X \sim B(N,p) \Leftrightarrow$$

$$X \sim B(N,p) \Leftrightarrow \Pr(X = x \mid N, p) = \binom{N}{x} p^x \cdot (1-p)^{N-x},$$

$$x = 0, 1, \cdots, N$$

Draft: Version 1

The Binomial Distribution

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

$$N! = N \cdot (N-1) \cdot (N-2) \cdot (N-3) \cdots 4 \cdot 3 \cdot 2 \cdot 1$$

- $\binom{N}{x}$: # of ways you can choose x from a group of N
- E[X] = N*p
- Var(X) = N*p*(1-p)

Draft: Version 1

DUAL RANDOM VARIABLE OF X:

X = "# Successes in a sequence of N trials",

Y = "# Failures in a sequence of N trials",

↓

Y = N-X

$$\Pr(Y = y \mid N, 1-p) = \Pr(N - X = y \mid N, p) = \Pr(X = N - y \mid N, p)$$

$$= \binom{N}{N-y} p^{N-y} \cdot (1-p)^{N-(N-y)} = \binom{N}{y} (1-p)^{y} \cdot p^{N-y}$$

Conclusion:

$$X \sim B(N,p) \Leftrightarrow Y \sim B(N,1-p)$$

Pretzel Example:

You are planning to sell a new pretzel and you want to know whether it will be a success or not. Initially, you are 50% certain that it will be a "Hit". Thus,

Pr("Hit") = Pr("Flop") = 0.5

If your pretzel is a **"HIT"** you expect to gain **30% of the market**. Let X be the number of people out of a group of N that buy your pretzel.

Pretzel Example: The Binomial Distribution

Assumption:

Pr(X=x| N, pretzel is a "Hit") = $\binom{N}{x} 0.3^x \cdot (0.7)^{N-x}$

Assumption:

Pr(X=x| N, pretzel is a "Flop") = $\binom{N}{x} 0.1^x \cdot (0.9)^{N-x}$

You decide to investigate the market for your pretzel and on a trial day it appeared that **5 OUT OF 20 PEOPLE** bought your pretzel.

What do you think now of your chances of the pretzel being a "Hit" or a "Flop"?

Draft: Version 1

Pretzel Example: The Binomial Distribution

Notation: Data = (20,5)

Calculation:

$$\Pr(''Hit'' \mid Data) =$$

$$\frac{\Pr(Data \mid ''Hit'')\Pr(''Hit'')}{\Pr(Data \mid ''Hit'')\Pr(''Hit'') + \Pr(Data \mid ''Flop'')\Pr(''Flop'')}$$

$$\Pr(Data \mid ''Hit'') = \binom{20}{5} 0.3^5 \cdot (0.7)^{15} = 0.179$$

(Table - Page 686)

Draft: Version 1

Pretzel Example: The Binomial Distribution

$$\Pr(Data\,|\,"Flop") = \binom{20}{5} 0.1^5 \cdot (0.9)^{15} = 0.032$$

(Table - Page 686)

Conclusion:

$$\Pr("Hit"\,|\,Data) = \frac{0.179 \cdot 0.5}{0.179 \cdot 0.5 + 0.032 \cdot 0.5} = 0.848$$

Further Development of selling Pretzel may be warranted!

The Poisson Process and Distribution

Consider a particular event
e.g. a customer arriving at a bank.

Assumptions:

1. The events can occur at any point in time.
2. The arrival rate per hour is constant, e.g.   customers per hour.
3. The number of customers arriving in disjoint time intervals are independent of each other, e.g. the number of customers in the first hour day and the number of customers in the second hour of the day.

Define:

X(t) = "# of such events in the time interval [0,t]"

Draft: Version 1

The Poisson Process and Distribution

- X(t) ~ Poisson(m*t):

$$\Pr(X = k \mid m, [0,t)) = \frac{(m \cdot t)^k}{k!} e^{-m \cdot t}$$

- $\Pr(Y = k \mid n) = \dfrac{n^k}{k!} \cdot e^{-n}$ is called the Poisson distribution
- E[Y] = n
- Var[Y] = n
- E[X] =m*t
- Var[X]= m*t

Pretzel Example: The Poisson Process

Based on your previous market research you decide to invest in a pretzel stand. Now you just need to select a good location. You consider your location to be **"good"**, **"bad"** or **"dismal"** if you sell **20**, **10** or **6**, respectively, per hour. You assume that customers arrive according to a Poisson Process.

X(t)="# of customer in the interval [0,t]"

- **Pr(X(t)=k|"Good")** = $\dfrac{(20 \cdot t)^k}{k!} e^{-20 \cdot t}$
- **Pr(X(t)=k|"Bad")** = $\dfrac{(10 \cdot t)^k}{k!} e^{-10 \cdot t}$
- **Pr(X(t)=k|"Dismal")** = $\dfrac{(6 \cdot t)^k}{k!} e^{-6 \cdot t}$

Draft: Version 1

Pretzel Example: The Poisson Process

Pr("GOOD") = 0.70, Pr("BAD") = 0.20, Pr("DISMAL") = 0.10

You give yourself one week for people to get to know you at this location. The second week, you open your stand in the morning and in the first half hour, 7 people bought your pretzel. Hmmm, You want to reevaluate your location.

SHOULD YOU RELOCATE?

Notation: Data = (7, [0,0.5))

We want to know:

$$\Pr("Good" \mid Data) = ?$$

Draft: Version 1

Pretzel Example: The Poisson Process

Calculation: $$\Pr(\text{"}Good\text{"} \mid Data) =$$

$$\frac{\Pr(Data \mid \text{"}Good\text{"})\Pr(\text{"}Good\text{"})}{\Pr(Data \mid \text{"}Good\text{"})\Pr(\text{"}Good\text{"}) + \Pr(Data \mid \text{"}Bad\text{"})\Pr(\text{"}Bad\text{"}) + \Pr(Data \mid \text{"}Dismal\text{"})\Pr(\text{"}Dismal\text{"})}$$

- Pr(Data|”Good”)=Pr(X(0.5)=7|”Good”) = $\frac{(20 \cdot 0.5)^7}{7!}e^{-20 \cdot 0.5}$

  = 0.09, **Table – Page 700.**
- Pr(Data|”Bad”)=Pr(X(0.5)=7|”Bad”) = $\frac{(10 \cdot 0.5)^7}{7!}e^{-10 \cdot 0.5}$

  = 0.104, **Table – Page 698**.
- Pr(Data|”Dismal”)=Pr(X(0.5)=7|”Dismal”) = $\frac{(6 \cdot 0.5)^7}{7!}e^{-6 \cdot 0.5}$

  = 0.022, **Table – Page 698.**

Draft: Version 1

Pretzel Example: The Poisson Process

- Pr(“Good”) = 0.70, Pr(“Bad”) = 0.20, Pr(“Dismal”) = 0.10

$$\Pr(\text{"}Good\text{"} \mid Data) = \frac{0.09 \cdot 0.70}{0.09 \cdot 0.70 + 0.104 \cdot 0.20 + 0.022 \cdot 0.10} = 0.733$$

- Simlarly: Pr(‘Bad”|Data) = 0.242, Pr(“Dismal|Data) = 0.025

Conclusion:

In light of the new data, you decide that the chances of this being a “Dismal” location for the pretzel stand is remote and your chance for this being a “Good” location has slightly improved. **You decide to stay.**

Draft: Version 1

The Exponential Distribution

Consider a particular event e.g. a customer arriving at a bank. Now consider, **the length of time between two consecutive events** e.g. the time between two customers arriving.

Alternative assumptions for Poisson process:

1. The arrival rate per hour is constant, e.g. m customers per hour.
2. Inter-arrival Times are exponentially distributed with parameter m.
3. Customers arrive independently from each other.

Define:

T = "Time between two consecutive customers arriving"

Draft: Version 1

The Exponential Distribution

- T ~ Exponential(m):

$$F_T(t \mid m) = \Pr(T \le t \mid m) = 1 - e^{-m \cdot t}$$

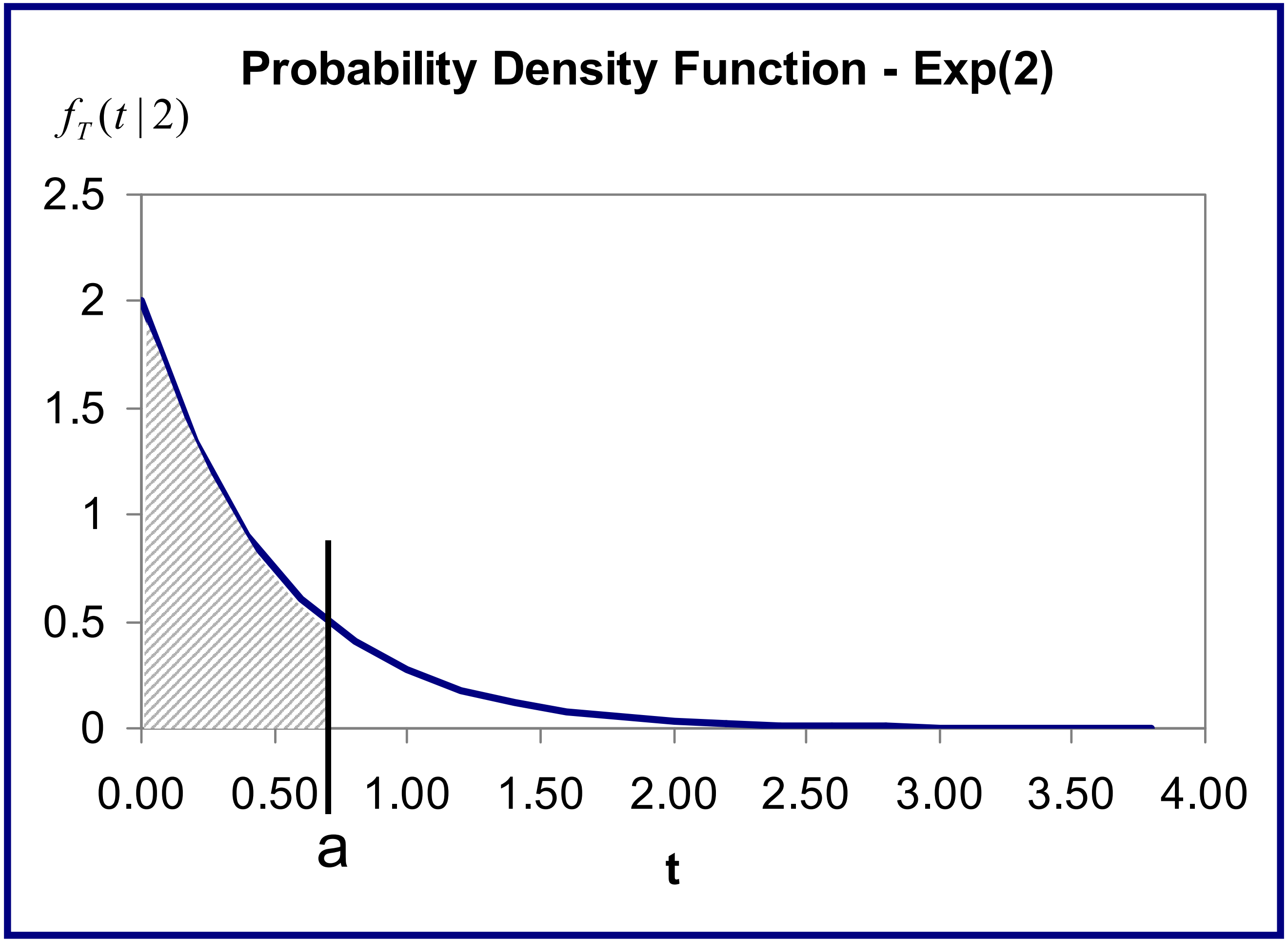
- $F_T(t \mid m):$ Cumulative Distribution Function of *T* (CDF)

- The density function follows from

$$f_T(t \mid m) = \frac{dF_T(t \mid m)}{dt} = m \cdot e^{-m \cdot t}$$

Draft: Version 1

The Exponential Distribution



The Exponential Distribution

- $\Pr(T \leq a \mid m) = 1 - e^{-m \cdot a}$

- $\Pr(T > a \mid m) = 1 - \Pr(T \leq a \mid m) = e^{-m \cdot a}$

- $\Pr(b < T \leq a \mid m) = \Pr(T \leq a \mid m) - \Pr(T \leq b \mid m) =$

$$1 - e^{-m \cdot a} - (1 - e^{-m \cdot b}) = e^{-m \cdot b} - e^{-m \cdot a}$$

- $E[T] = \dfrac{1}{m}, Var(T) = \dfrac{1}{m^2}$

Draft: Version 1

Pretzel Example: Exponential Distribution

You want to provide fast service for your customers and you are wandering whether you can in your current setup of your stand. It takes approximately **3.5 minutes to cook a pretzel.** What is the **probability that the next customer arrives before the pretzel is finished**. You recall your initials assumptions, i.e. You assume that customers arrive according to a Poisson Process and you consider your location "good", "bad" or "dismal" if you sell 20, 10 or 6, respectively, per hour. Initially, you belief that for your first location:

Pr("Good") = 0.70, Pr("Bad") = 0.20, Pr("Dismal") = 0.10

Pretzel Example: Exponential Distribution

Calculation:

$$\begin{aligned}\Pr(T > 3.5Min) = {} & \Pr(T > 3.5Min \,|\, \text{"}Good\text{"}) \cdot \Pr(\text{"}Good\text{"}) + \\ & \Pr(T > 3.5Min \,|\, \text{"}Bad\text{"}) \cdot \Pr(\text{"}Bad\text{"}) + \\ & \Pr(T > 3.5Min \,|\, \text{"}Dismal\text{"}) \cdot \Pr(\text{"}Dismal\text{"})\end{aligned}$$

$$\begin{aligned}\Pr(T > 3.5Min \,|\, \text{"}Good\text{"}) &= \Pr(T > \frac{3.5}{60} hours \,|\, 20) = \\ & \Pr(T > 0.0583 \,|\, 20) = e^{-20 \cdot 0.0583} = 0.3114\end{aligned}$$

$$\begin{aligned}\Pr(T > 3.5Min \,|\, \text{"}Bad\text{"}) &= \Pr(T > \frac{3.5}{60} hours \,|\, 10) = \\ & \Pr(T > 0.0583 \,|\, 10) = e^{-10 \cdot 0.0583} = 0.5580\end{aligned}$$

Pretzel Example: Exponential Distribution

$$\Pr(T > 3.5Min \mid "Dismal") = \Pr(T > \frac{3.5}{60} hours \mid 6) =$$
$$\Pr(T > 0.0583 \mid 6) = e^{-6 \cdot 0.0583} = 0.7047$$

Pr(“Good”) = 0.70, Pr(“Bad”) = 0.20, Pr(“Dismal”) = 0.10

$$\Pr(T > 3.5Min) = 0.3114 \cdot 0.7 + 0.5580 \cdot 0.2$$
$$+ 0.7047 \cdot 0.1 = 0.40$$

Or in other words: $\Pr(T \leq 3.5Min) = 0.60$

Conclusion:

60% of your customers will have to wait until the pretzel is ready.

Pretzel Example: Exponential Distribution

You realize that customers prefer hot pretzels and you are not to concerned about this number. However you decide to reevaluate after one week of operation. The second week, you open your stand in the morning and in the first half an hour 7 people brought your pretzel. What do you think now of is the percentage of people waiting for a pretzel.

Notation: Data = (7, [0,0.5))

Pretzel Example: Exponential Distribution

$$\Pr(T > 3.5Min \mid Data) =$$

$$\Pr(T > 3.5Min \mid "Good", Data) \cdot \Pr("Good" \mid Data) +$$
$$\Pr(T > 3.5Min \mid "Bad", Data) \cdot \Pr("Bad" \mid Data) +$$
$$\Pr(T > 3.5Min \mid "Dismal", Data) \cdot \Pr("Dismal" \mid Data)$$

$$\Pr(T > 3.5Min \mid "Good", Data) =$$
$$\Pr(T > 3.5Min \mid "Good") = 0.3114$$

$$\Pr(T > 3.5Min \mid "Bad", Data) =$$
$$\Pr(T > 3.5Min \mid "Bad") = 0.5580$$

$$\Pr(T > 3.5Min \mid "Dismal", Data) =$$
$$\Pr(T > 3.5Min \mid "Dismal") = 0.7047$$

Draft: Version 1

Pretzel Example: Exponential Distribution

$$\Pr(\text{"}Good\text{"} \mid Data) = 0.733 \qquad \Pr(\text{"}Bad\text{"} \mid Data) = 0.242$$

$$\Pr(\text{"}Dismal\text{"} \mid Data) = 0.025$$

Hence:

$$\Pr(T > 3.5Min \mid Data) =$$

$$0.3114 \cdot 0.733 + 0.5580 \cdot 0.242 + 0.7047 \cdot 0.025 = 0.3809$$

Or in other words: $\Pr(T \leq 3.5Min \mid Data) = 0.62$

Conclusion:

62% of your customers will have to wait until the pretzel is ready (which increased from the previous 60%. You are concerned about the chance of customers waiting increasing. You decide to continue to monitor this percentage and may consider investing in another pretzel oven.

Draft: Version 1

The Normal Distribution

Consider the production of men shoes. You want to offer these shoes in many different sizes. However, you need to decide the percentage of shoes to produce in each size. Let *Y* be the length of men's feet. Many biological phenomena (height, weight, length) follow a **bell-shaped curve** that can be represented by a normal distribution.

- $Y \sim N(\mu, \sigma)$:

$$f_Y(y \mid \mu, \sigma) = \frac{1}{\sigma\sqrt{2 \cdot \pi}} \cdot e^{-\frac{(x-u)^2}{2\sigma^2}}$$

- $E[Y] = \mu$
- $Var(Y) = \sigma^2$

Draft: Version 1

The Normal Distribution

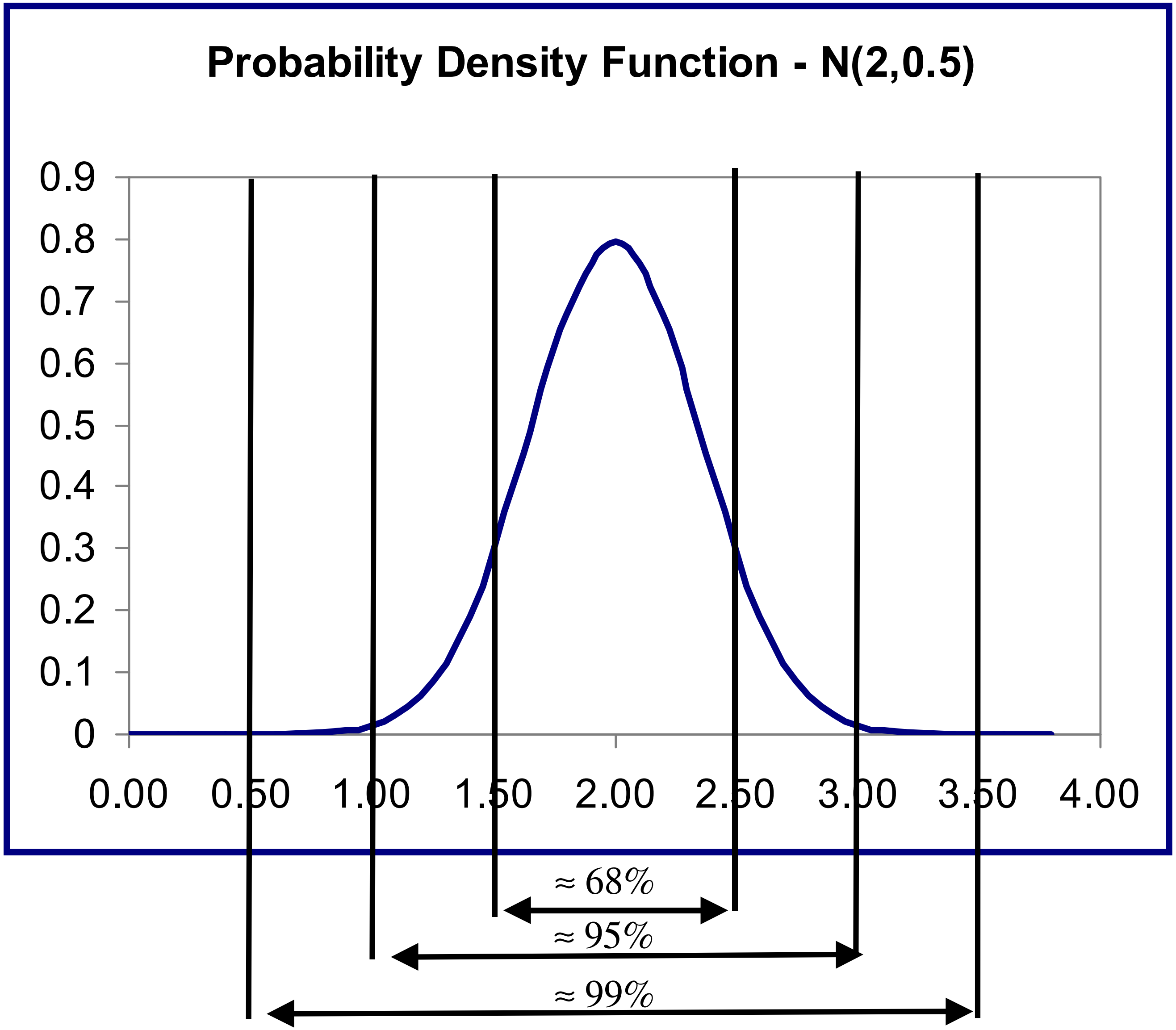
- Some handy rules of thumb:

$$\Pr(\mu - \sigma < Y < \mu + \sigma) \approx 0.68$$

$$\Pr(\mu - 2\sigma < Y < \mu + 2\sigma) \approx 0.95$$

$$\Pr(\mu - 3\sigma < Y < \mu + 3\sigma) \approx 0.99$$

The Normal Distribution



The Normal Distribution

Define:

$$Z = \frac{Y - \mu}{\sigma}$$

- *Z* Standard Normal Distributed $\Leftrightarrow$ $Z \sim N(0,1)$
- The Standard Normal CDF is available in Table Format.
- Normal distribution is symmetric around its mean:

$$\Pr(Y - \mu < -y \mid \mu, \sigma) = \Pr(Y - \mu > y \mid \mu, \sigma) \Leftrightarrow$$

$$\Pr(Z < -z) = \Pr(Z > z)$$

The Normal Distribution

How do we calculate $\Pr(a < Y \le b \mid \mu, \sigma)$
if only the CDF for Z is available in Table format?

Convert to a Standard Normal Distribution:

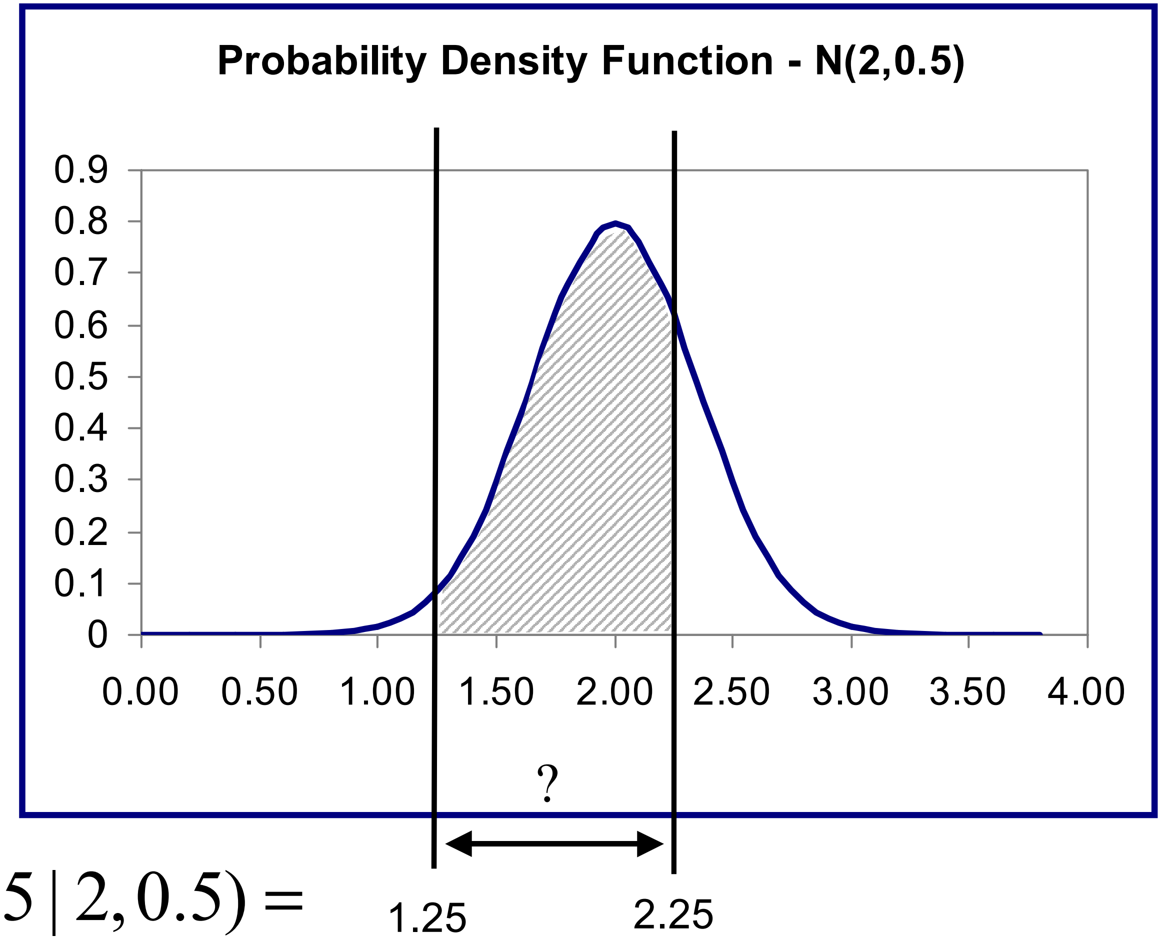
$$\Pr(a < Y \le b \mid \mu, \sigma) = \Pr\left(\frac{a-\mu}{\sigma} < \frac{Y-\mu}{\sigma} \le \frac{b-\mu}{\sigma} \mid \mu, \sigma\right) =$$

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

Draft: Version 1

The Normal Distribution

Example:

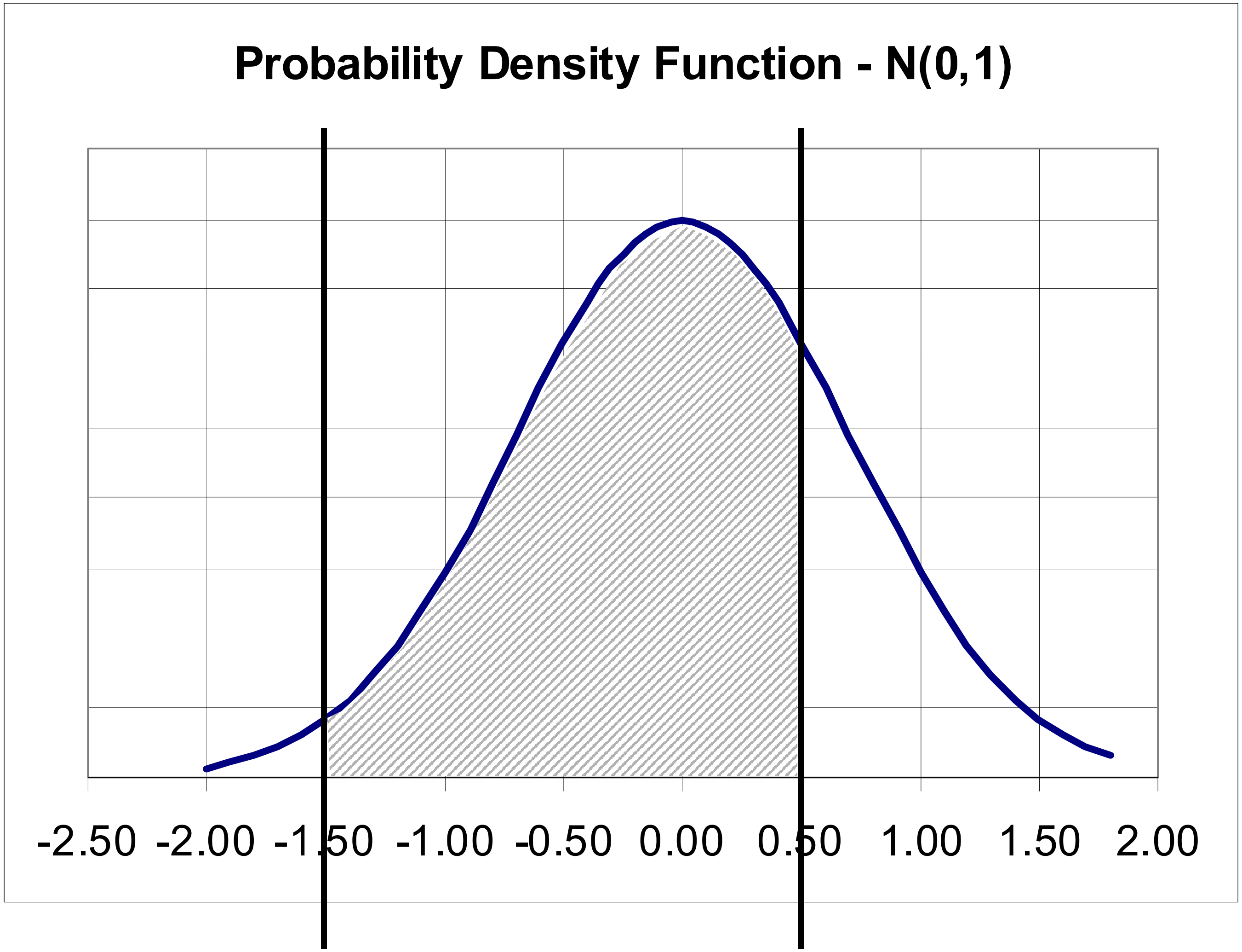


$$\Pr(1.25 < Y \leq 2.25 \mid 2, 0.5) =$$

$$\Pr\left(\frac{1.25-2}{0.5} < Z \leq \frac{2.25-2}{0.5}\right) = \Pr\left(-\frac{3}{2} < Z \leq \frac{1}{2}\right)$$

Draft: Version 1

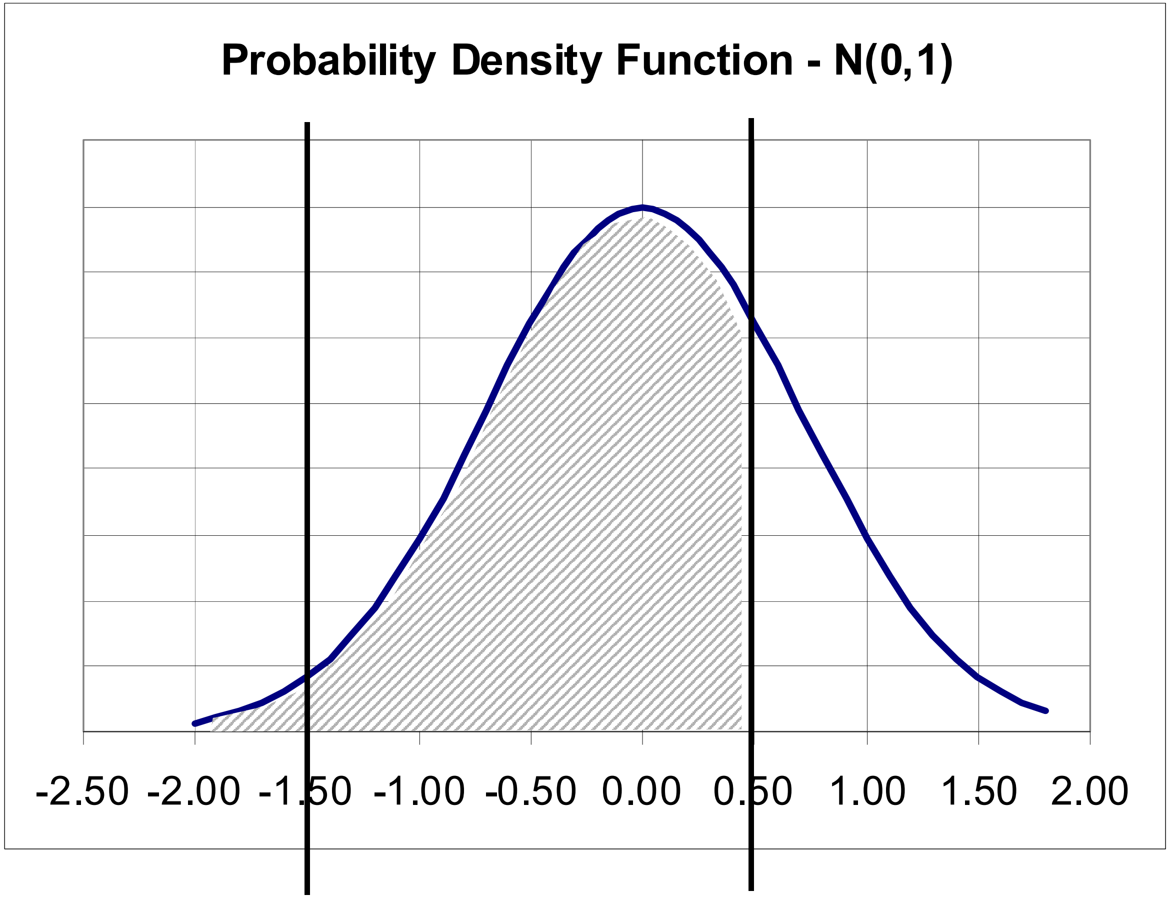
The Normal Distribution



$$\frac{1.25-2}{0.5} \approx -1.50 \qquad \frac{2.25-2}{0.5} \approx 0.50$$

Draft: Version 1

The Normal Distribution



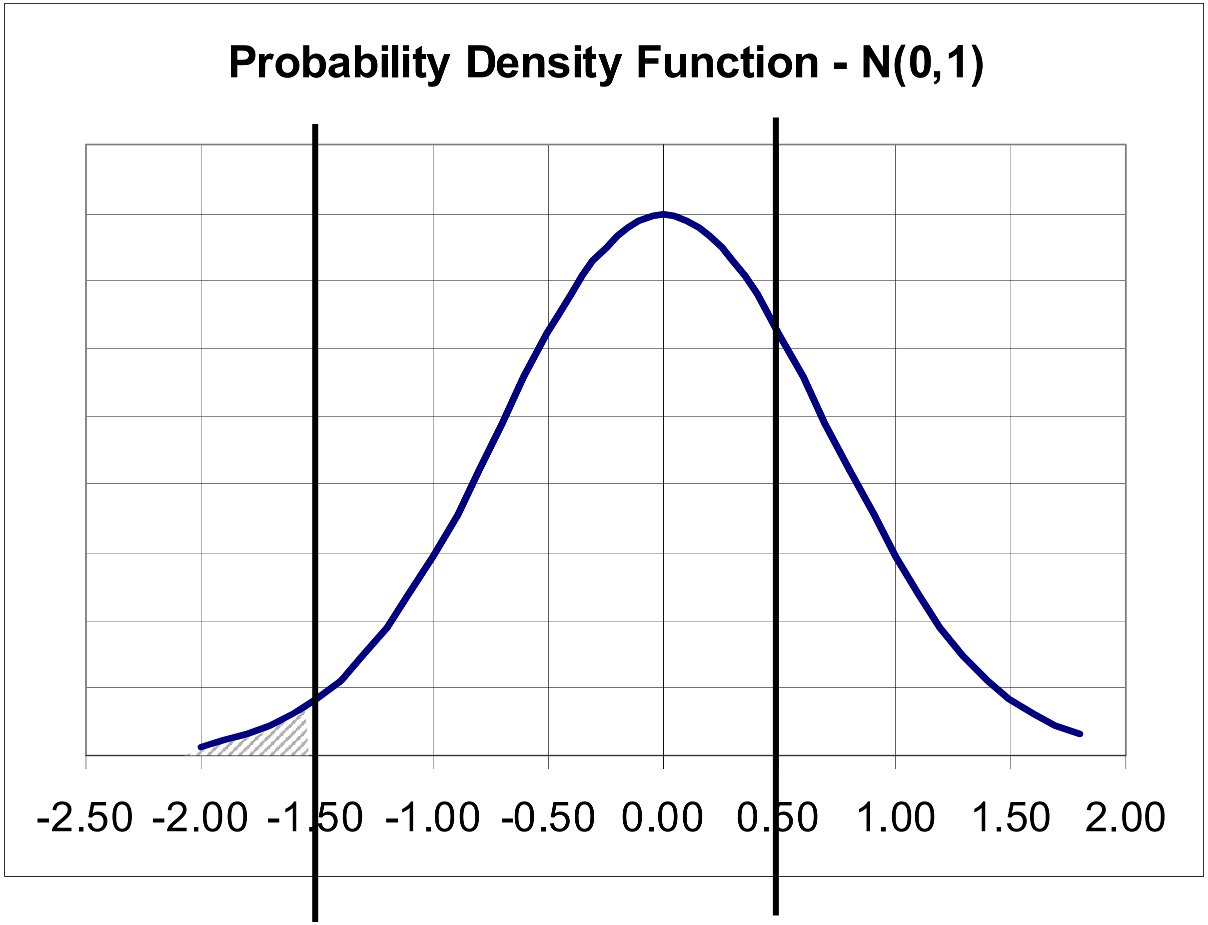
$$\frac{1.25-2}{0.5} \approx -1.50 \qquad \frac{2.25-2}{0.5} \approx 0.50$$

$$\Pr(Z \leq \frac{1}{2}) = 0.6915$$

See Table – Page 708

Draft: Version 1

The Normal Distribution



$$\frac{1.25-2}{0.5} \approx -1.50 \qquad \frac{2.25-2}{0.5} \approx 0.50$$

$$\Pr(Z \le -\frac{3}{2}) = 0.0668$$

See Table – Page 707

Draft: Version 1

Conclusion:

$$\Pr(-\frac{3}{2} < Z \le \frac{1}{2}) = \Pr(Z \le \frac{1}{2}) - \Pr(Z \le -\frac{3}{2})$$

$$= 0.6915 - 0.0668 = 0.6247$$

QUALITY CONTROL EXAMPLE:

You are the producer of hard drives for personal computers. One of your machines that produces a part is used in the final assembly of the disk drive. The width of this part is important for the proper functioning of the hard drive. If the width **falls below 3.995mm** or the width **falls above 4.005mm**, the hard drive will not function properly. If the disk drive does not work, it must be repaired at a cost of $10.40.

QC Example: The Normal Distribution

The machine can be set a width of 4mm, but it is not perfectly accurate. The production speed of the machine can be set high or low. However, **the higher production speed result in lower accuracy**. In fact, if W is the width of the part:

(W| High Production Speed) ~ N(4, 0.0026)

(W| Low Production Speed) ~ N(4, 0.0019)

Of course at a higher production speed more hard drives are produced and the **cost per hard drive is $20.45.** At the lower production speed the **cost per hard drive is $20.75.**

Should you turn at high production speed or low production speed?

Draft: Version 1

QC Example: The Normal Distribution

Calculation: Production At Low Speed

$$\Pr(\text{Defective} \mid \text{Low Speed}) = 1 - \Pr(\text{Not Defective} \mid \text{Low Speed})$$

$$= 1 - \Pr(3.995 < W \leq 4.005 \mid \mu = 4, \sigma = 0.0019) =$$

$$1 - \Pr\left(\frac{3.995 - 4}{0.0019} < \frac{W - 4}{0.0019} \leq \frac{4.005 - 4}{0.0019} \mid \mu = 4, \sigma = 0.0019\right)$$

$$= 1 - \Pr(-2.63 < Z \leq 2.63) = 1 - \left(\Pr(Z \leq 2.63) - \Pr(Z \leq -2.63)\right)$$

$$1 - (0.9957 - 0.0043) = 1 - 0.9914 = 0.0086$$

Draft: Version 1

See Table – Page 709,707

QC Example: The Normal Distribution

Calculation: Production at High Speed

$$\Pr(\text{Defective}|\text{ High Speed}) = 1 - \Pr(\text{Not Defective}|\text{ High Speed})$$

$$= 1 - \Pr(3.995 < W \leq 4.005 \mid \mu = 4, \sigma = 0.0026) =$$

$$1 - \Pr\left(\frac{3.995 - 4}{0.0026} < \frac{W - 4}{0.0026} \leq \frac{4.005 - 4}{0.0026} \mid \mu = 4, \sigma = 0.0026\right)$$

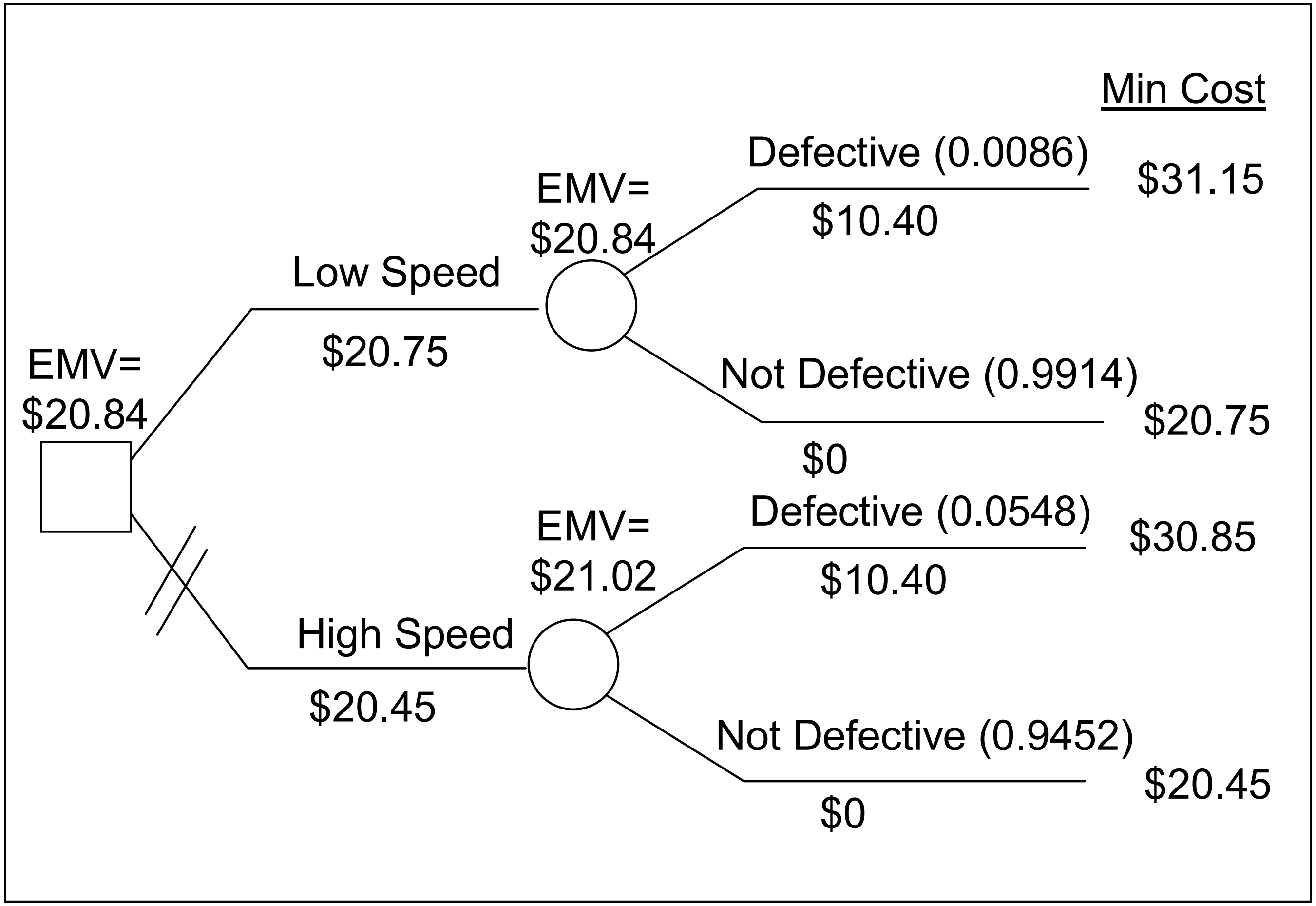
$$= 1 - \Pr(-1.92 < Z \leq 1.92) = 1 - \left(\Pr(Z \leq 1.92) - \Pr(Z \leq -1.92)\right)$$

$$1 - (0.9726 - 0.0274) = 1 - 0.9452 = 0.0548$$

See Table – Page 709,707

Draft: Version 1

QC Example: The Normal Distribution



Conclusion: Run at a slower speed. Increased cost from slow speedare offset by the increased precision.

Draft: Version 1

The Beta Distribution

Suppose you are interested in the proportion of voters in your town that will vote for the next republican president. This proportion is uncertain and may range from 0 to 1.

Let Q be that proportion and assume Q~Beta(n,p)

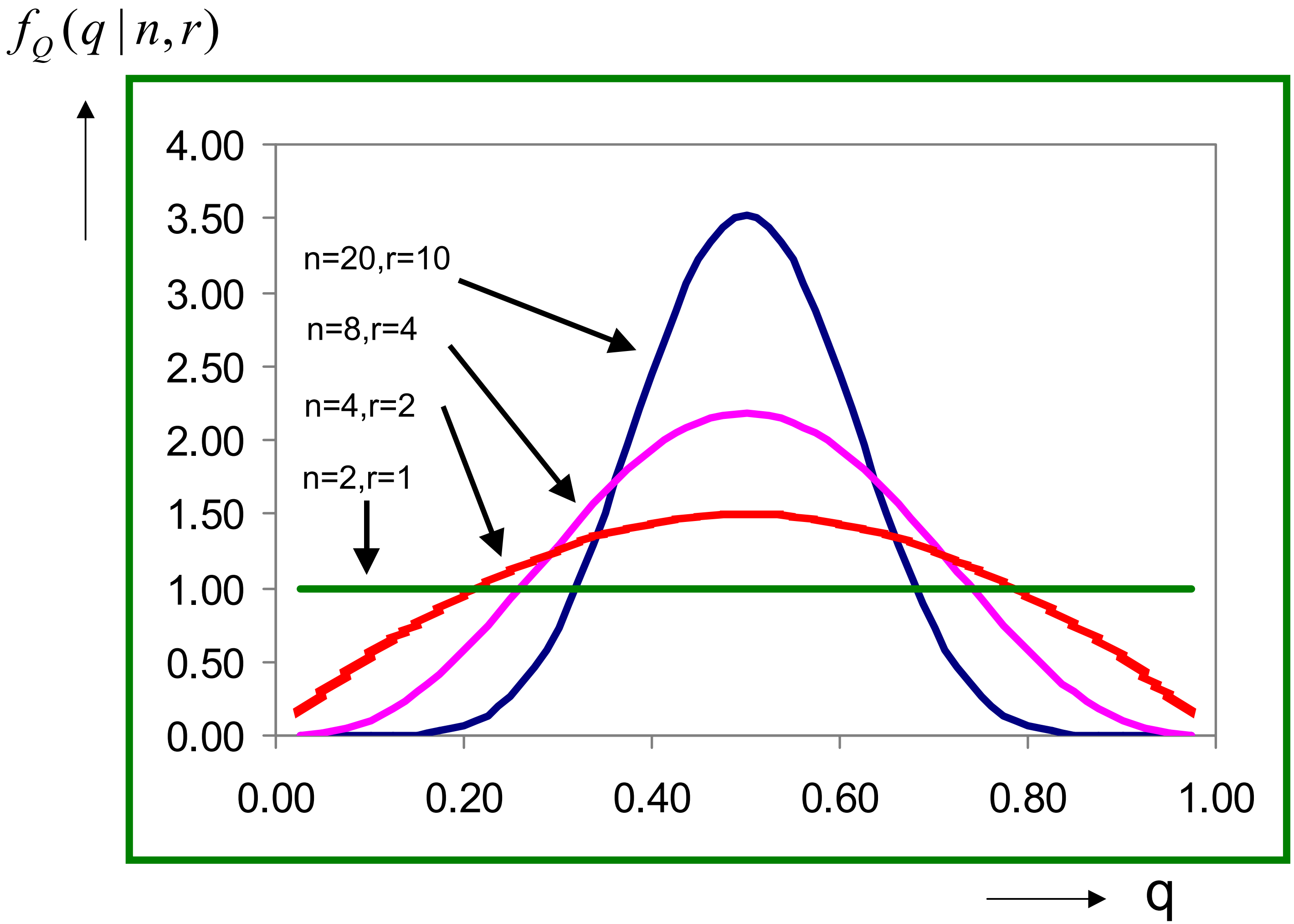
$$f_Q(q \mid n,r) = \frac{\Gamma(n)}{\Gamma(r)\cdot\Gamma(n-r)} \cdot q^{r-1}(1-q)^{n-r-1}, 0 < q < 1$$

$$\Gamma(n) = (n-1)! = (n-1)\cdot(n-2)\cdots 3\cdot 2\cdot 1, n = 1,2,3,...$$

Draft: Version 1

The Beta Distribution

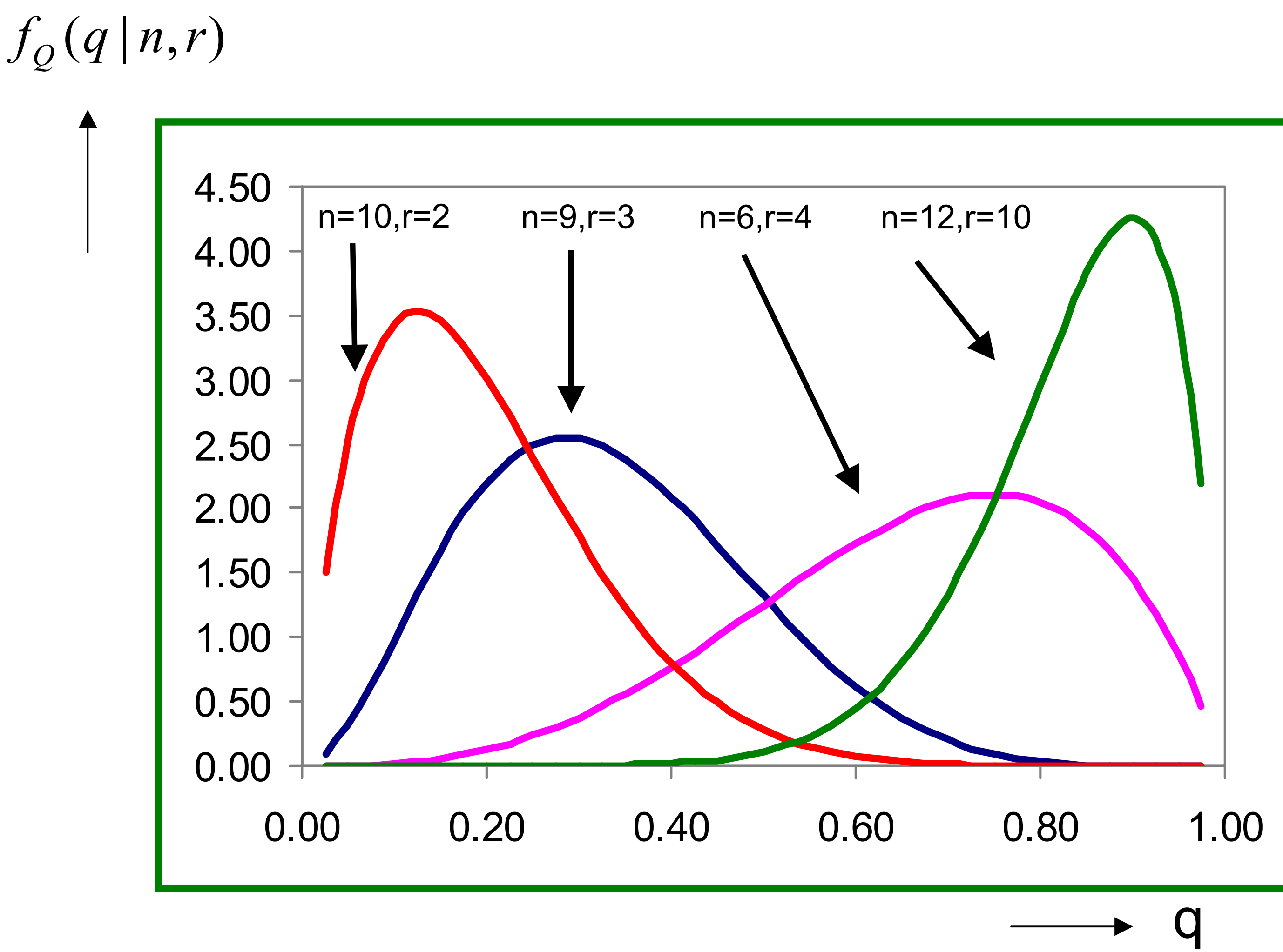
SYMMETRIC BETA DISTRIBUTIONS



Draft: Version 1

The Beta Distribution

ASYMMETRIC BETA DISTRIBUTIONS



Draft: Version 1

The Beta Distribution

$$E[Q] = \frac{r}{n} \qquad Var(Q) = \frac{r(n-r)}{n^2(n+1)}$$

Elicitation Of Parameters Using Informal Parameter Interpretation:

n = “Number of Trials”
r = “Number of Successes”

EXAMPLE:
You first guess for the preference of the Republican Candidate is that **4 out of 10** people would vote for the Republican Candidate. You set: **n=10, r=4.**

Note this coincides with **an expected proportion of 40%.**

Draft: Version 1

The Beta Distribution

After talking to people on the street you **reevaluate** your beliefs and estimate that **40 out of 100** people would vote for the Republican Candidate. You set: **n = 100, r = 40.** Note that this still also coincides with an **expected proportion of 40%.**

What is the difference with the previous estimate?

First Estimate:

$$St.Dev.(Q) = \sqrt{\frac{4(10-4)}{10^2(10+1)}} = 14.7\%$$

Second Estimate:

$$St.Dev.(Q) = \sqrt{\frac{40(100-40)}{100^2(100+1)}} = 4.9\%$$

Draft: Version 1

Pretzel Example: The Beta Distribution

You want to re evaluate your decision to invest in a pretzel stand. Sales have been okay in the first week, but not too great. You are wandering whether you should proceed. You estimate at this point that you are **50% sure that your market share is less than 20%** and your **75% sure that your market share is less than 38%.**

Let *Q* be the proportion of the market. You decide to model your uncertainty in *Q* as **a beta distribution** and using the table on **page 711** that:

$$\Pr(Q \leq 0.20 \mid n = 4, r = 1) = 0.49$$

$$\Pr(Q \leq 0.38 \mid n = 4, r = 1) = 0.76$$

Pretzel Example: The Beta Distribution

You decide that that is close enough and proceed with the analysis. You estimate that the **total monthly market is 100,000 pretzels**. Your price for a pretzel is set at **$0.50** and it costs you **$0.10** to produce a pretzel. You estimate **$8000 of monthy fixed cost** for your pretzel stand and some overhead. Given the market share Q, you calculate for your net monthly profit:

$$\text{Net Profit} = \text{Revenue} - \text{Cost}$$
$$100000 * Q * \$0.50 - (100000 * Q * \$0.10 + 8000)$$
$$= 40000 * Q - 8000$$

However, **Q is uncertain** so you decide to calculate your **expected profit.**

Pretzel Example: The Beta Distribution

E[Profit] = E[40000*Q-8000] = 40000*E[Q]– 8000

$$E[Q] = \frac{r}{n} = \frac{1}{4} = 25\% \Rightarrow$$

$$E[\text{Profit}] = 40000 \times \frac{1}{4} - 8000 = \$2000$$

You start to be **more comfortable** with your decision to start a pretzel career, but **careful as you are**, you decide to evaluate your chances of loosing money.

$\Pr(\text{Net Profit} \le 0) = \Pr(Q \le 0.20 | n=4, r=1) = 0.49$

See Table Page 711

Conclusion: There is approximately 50% chance of loosing money. **Are you willing to continue to take this RISK?**

Draft: Version 1