Chapter 1

Introduction

1.1 Telecommunications and Quality Control

A book on applied probability would be pointless without applications. There are a huge number of possibilities including applications to the biological sciences, to manufacturing or to the behavioural sciences but here we emphasize applications to telecommunications and to quality control.

1.2 The Digital Network

Real world applications of applied probability are as near at hand as your telephone. The information revolution is upon us. The integration of computing and telecommunications will change the way people live and work. Traditional services such as mail, bank transactions and newspaper subscriptions will be delivered electronically directly to the home along with telephone services. New services will emerge that we can't imagine. These services will be delivered on the information highway built on a network of fiber optical cables.

The traffic laws for this highway are hotly debated. Asynchronous transfer mode or ATM was conceived in the late nineteen eighties as international standard for the integrated services digital networks or ISDN networks capable of carrying the above mixture of services. This standard was designed to deliver the quality of service we expect from a telephone network. However, the advent of the world wide web changed everything! The light and easy internet protocol TCP/IP (Transfer Control Protocol over the Internet Protocol) was better adapted for delivering web pages. Today the TCP/IP protocol dominates but so far falls short in delivering the quality of service envisaged for ATM.

Both protocols are based on sending information in packets across the network. Under both ATM and TCP/IP, multimedia services such as digitized voice, text, image, video and computer communications are supported by dividing the data stream into ATM cells or TCP/IP packets. ATM cells are short, 53 byte packets while TCP/IP packets are of different sizes. These heterogeneous data streams can be multiplexed together over a common transmission medium such as an optical cable. Consequently this high capacity medium must no longer be dedicated to a single data source.

In both protocols the routing information is in the header and the data follows. The format of an ATM cell is given in Figure 1.1 below:

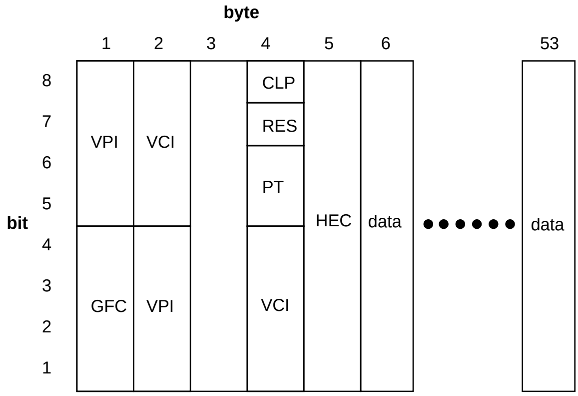

Fig. 1.1 The ATM cell

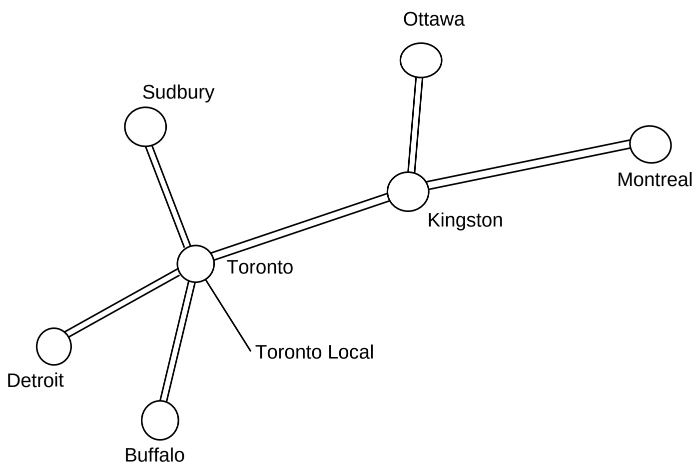

Fig. 1.2 An ATM network

On the information highway, asphalt and cement are replaced by fiber optical cables. Motor vehicles become cells or packets and the highway interchange is replaced by an electronic ATM switch or packet router. If vehicle traffic obeyed the laws of an ATM network then all vehicles would be of the same size and capacity and they would carry passengers or cargo. A big delivery from Toronto to Ottawa could consist of a series or convoy of vehicles merged or multiplexed onto the high capacity 401 highway leading to Kingston along with all other vehicles headed that way. The drivers of the vehicles wouldn't know the destination but would carry an identifier which a dispatcher in Kingston would recognize. The dispatcher at the Kingston interchange would be looking for vehicles with this identifier and would know that these vehicles should be switched onto highway 15 toward Ottawa. There is in fact a virtual connection for all the vehicles involved in this delivery maintained by the dispatchers along the path. In an ATM cell in Figure 1.1 the passengers or cargo are the data in the 48 byte data field. The cell identifier is given in fields VPI (Virtual Path Indicator) and VCI (Virtual Channel Indicator).

If vehicle traffic obeyed the laws of a TCP/IP network then vehicles would come in variable sizes and capacities. A big delivery from Toronto to Ottawa would consist of a convoy of vehicles with drivers who know their final destination. When the vehicles arrive at the Kingston interchange the driver would tell the dispatcher his destination and the dispatcher would look up the best road in a routing table. The vehicle would then merge onto highway 15 if directed to do so. There is no virtual connection so in this sense an TCP/IP network is more like a real vehicle highway than an ATM network. TCP/IP is light and easy with no prearranged virtual path but as with the real traffic network there is no way to limit access to avoid traffic jambs! This is the first major difference between ATM and TCP/IP. Finally, under TCP/IP, when each vehicle arrives Ottawa a small vehicle is dispatched back to Toronto acknowledging that this portion of the delivery was successfully made. This acknowledgement feedback in TCP/IP is another major difference between ATM and TCP/IP.

Let's consider what happens in an ATM network when a long distance telephone call is made from Toronto to Ottawa. When the number is dialed the signalling system called SS7 must set up the call. SS7 alerts the ATM switches in Toronto, Kingston and Ottawa that it has a call requiring a capacity of 64,000 bits a second plus the same amount to carry the other party's voice back to the caller. If that capacity is not available then the caller gets a busy signal. If the call is accepted SS7 ties up the resources and rings the other party. If the other party doesn't answer and the caller hangs up then SS7 will release the resources. If the other party answers, the ATM switches are notified to expect cells with given VCI-VPI identifiers. The Toronto switch knows it must send cells with this identifier to Kingston while the Kingston switch knows it must send such cells to Ottawa (not Montreal for instance). The Ottawa switch knows it sends cells with this identifier to a specific telephone number in Ottawa. This completes the ATM virtual circuit.

The mechanics of the same call on a TCP/IP network are still in flux. When conceived in 1983, TCP/IP was designed to deliver packets like registered letters sent through the postal service. Packets pass through a series of routers, are stored and then sorted and then sent on their way. When a letter is successfully delivered an acknowledgement is returned to the source. This protocol was designed for reliable delivery over a failure prone network with military applications in mind. The concept of virtual circuits is foreign to TCP/IP and this prevents telephone operators from guaranteeing high quality calls. Nevertheless by tweaking TCP/IP, voice over TCP/IP is becoming common and this is pushing down the cost of personal communication.

For simplicity we will mostly consider ATM here since switching fixed length cells is easier to analyze than routing variable length packets (in fact some TCP/IP routers break packets in fixed length cells for switching purposes and then reassemble the packet at the output port). One should bear in mind that technological changes like optical switching will eventually make both these protocols obsolete. Nevertheless, if past history is any guide, the mathematical concepts studied here will remain relevant for evaluating the performance of newer technologies.

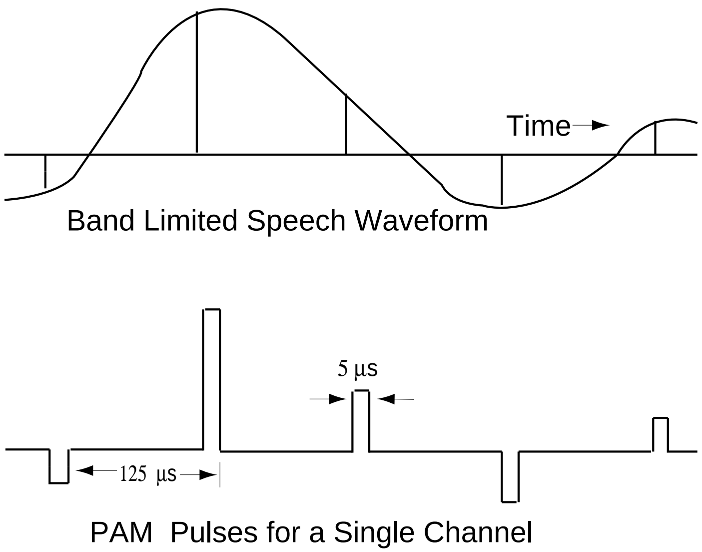

Fig. 1.3 Speech sampling

The mechanics of converting voice to information hasn't changed in 40 years. The caller's voice compresses a diaphragm in the mouth piece of the telephone which generates an electric current which is detected at the local telephone wire center. This signal voltage is measured or sampled at the wire center every 125 microseconds or 8,000 times a second. The voltage is quantized and encoded into 8 bits or 1 byte; that is 64,000 bits or 8000 bytes are produced every second. Six milliseconds worth of speech makes 48 bytes which is exactly the data content of an ATM cell. This is about right because the human ear can't distinguish delays shorter than 6 milliseconds. Similarly TCP/IP packets may carry about 6 milliseconds worth of voice data from a single source. Hence each speaker would produce a steam of cells or packets at a rate of about 6 per millisecond.

Alternatively the cell may serve as a frame to carry one byte from 48 different speakers between the same two points on the network. If 48 customers were simultaneously calling Ottawa from Toronto they each would produce a byte every 1/8000 of a second and these bytes would be assembled into one cell and sent. At the receiving end these bytes would be disassembled and send to the separate receivers. This would produce a stream of cells at a rate of 8000 per second. In this way ATM can emulate the TDM (Time Division Multiplexing) system currently in use around the world. Naturally more customers can be accommodated by managing separate streams of frames. The same techniques can be used to carry voice over IP (VoIP).

The cells associated with our speaker arrive in Kingston, the header is identified and the ATM switch in Kingston switches these cells onto the Ottawa trunk. On arrival in Ottawa the header is again identified and the switch routes these cells to the local wire center in Ottawa. At the local wire center the digitization procedure is reversed and a voltage is sent down the line to the other party. This voltage drives the speaker in the other party's telephone and the speaker's words are received.

The projected rate of the trunk lines is 10 gigabytes per second; that is around 10,000,000,000 bytes per second or around 200,000,000 cells per second. On an ATM network this means that between consecutive cells from our speaker the switch sends out roughly 1,000,000 cells. These cells could be used for other speakers so in principal a million other callers could be using this trunk simultaneously! The enormous trunk capacity will of course be used for other kinds of traffic. Combining the cells or packets of various sources onto a single trunk is called multiplexing and this multiplexing of traffic results in a substantial increase in the carrying capacity of a single trunk. In fact since no source books a particular time slot, if a source has no cell or packet to send in a given time slot then this spare capacity can be used by somebody else. This is exactly what occurs in a telephone conversation. There are silence periods when the other party is talking. This means in fact that several million callers could use the same trunk line because most of any conversation is silence. Of course if everyone were speaking simultaneously and generating cells or packet at the peak rate then the trunk capacity will be inadequate. We hope this has small probability!

This is the basis of statistical multiplexing. More sources are accepted than could be handled if all sources transmit at peak rate and this increases the revenues of the telephone company. There is however an inevitable cost to pay. Conflicts or contention for resources will arise. A videoconference will generate about a million cells a second so a service of this kind will occupy a nonnegligable part of the trunk capacity. Hence the question of call admission becomes critical because too many videoconferences would disrupt telephone service to thousands of customers. A large computer file transfer might have the same effect but since a data transfer is not usually time critical such a call might be accepted but delayed at a switch if that switch gets too busy. Telephone calls are very delay sensitive however so they must get high priority. On the other hand if the switch is really overloaded it might simply drop a few cells associated with a telephone call and nobody would ever notice. Dropping a few cells in a data transfer would would be costly however, as the correct data might have to be retransmitted. Dropping the formatting cells in a video message might produce nonsense at the other end.

A simple large switch or router has no queueing at the input ports. Cells or packets are routed through the switch without queueing delay directly to buffers (or to a common buffer) at the appropriate output port. The cells or packets are then scheduled for transmission along the output link. In Exercise 1.1 we consider one output port comprising two buffers and a link in a 2×2 ATM switch. The point of the exercise to investigate the impact of different scheduling protocols for cells queued at the two competing output buffers. In fact modern switches now run more sophisticated protocols designed not only to reduce queue lengths and the associated delay but also to reduce the variability of the queueing delay so cells eventually arrive at the destination in a steady predictable stream.

The net effect of multiplexing many streams of data through a switch is clearly enormously complicated. It must be understood however because it is essential to decide ahead of time just how many traffic sources can be routed through this switch in order to avoid unacceptable delays and cell losses. This problem of admission control is still being hotly disputed. The problem is further complicated by the fact that a switch is receiving traffic directly from local sources or even local networks and also from other switches.

The performance of the ATM switch or the TCP/IP router will be judged not only on the average amount of traffic carried. It is also important to predict the proportion of cells or packets dropped as well as the average delay and the cell delay variation of cells or packets traversing the network . This results in complicated (but interesting) problems in *queueing theory*: the mathematical (and often probabilistic) theory of queues.

1.3 Quality Control

Many maintain the catastrophic decline of the North American automobile industry in the nineteen seventies and eighties resulted partly from the fact that foreign competitors adopted military quality control standards while the North American companies forgot all about quality. Quality control has many aspects. Acceptance sampling described below can be used by a buyer to force a supplier to deliver product of a specified quality. The supplier can avoid poor quality product by monitoring the production line using the on-line quality control schemes described in future chapters. The supplier can also design his product in such a way that minor imperfections in production do not result a poor quality product. The search for these robust production regimes is called off-line quality control and is not treated in this book.

The most famous military acceptance standards are MIL-STD-414 for acceptance sampling by variables and MIL-STD-105D for acceptance sampling by attributes. The former is used when a quality measurement is available for elements of the sample. The later is used when one can only determine if the elements of the sample are defective or not. We consider MIL-STD-105D for historical reasons since MIL-STD-105D was replaced by MIL-STD-105E in 1989 and then by MIL-STD-1916 in 2001. MIL-STD-105D has also been incorporated into the International Standards Organization (ISO) standard called ISO 2859. Department of Defence (DOD) Specifications and Standards are available for public use through the DOD Scientific and Technical Information Network at the http://stinet.dtic.mil web site.

Essentially a standard is used like a contract between a supplier and a buyer. The two sides agree on a price and both sides agree the buyer will accept the product according to the procedures carefully set out in the standard. These procedures essentially punish the supplier if he produces an unreasonably high proportion of defective or nonconforming units. On the other hand he is rewarded if he produces a reasonably low proportion of nonconforming units. The key word is reasonable, and this is spelled out by the concept of acceptable quality level - AQL.

The AQL, agreed to contractually by the supplier and the buyer, is the percentage of nonconforming units in lots that will be accepted most of the time by the sampling scheme. In other words, if the lots submitted have a percentage of nonconforming units no greater than the AQL then the sampling scheme will accept the great majority of these lots. In practice the great majority means about 95%. The standard does caution however that this does not give the supplier the right to knowingly supply any nonconforming unit of product!

Imagine that a supplier produces resistors in large batches and a buyer wants to sign a long term contract for one lot of resistors every working day. The first step is to agree on an AQL. The buyer would like an AQL of 0 of course but the supplier knows he can't meet that standard at least not at a reasonable price. They settle on an AQL of 2.5% since the buyer knows he can easily detect the defective resistors in the process of building his product and therefore he is willing to do the necessary screening for a lower price.

They next agree that all lots will contain 1000 units. This determines the sample size code letter. Therefore the sample size code letter J is picked from the table below.

Lot size			General inspection level II
2	to	8	A
9	to	15	B
16	to	25	C
26	to	50	D
51	to	90	E
91	to	150	F
151	to	280	G
281	to	500	H
501	to	1,200	J
1,201	to	3,200	K
3,201	to	10,000	L
10,001	to	35,000	M
35,001	to	150,000	N
150,001	to	500,000	Q
500001	and	over	R

Next look at Figure 1.4. This outline describes four regimes for the scheme. The usual regime is normal inspection. If the supplier delivers very good quality a level of trust is established and the regime of reduced inspection is entered. This reduces the cost of sampling to the buyer. If the supplier delivers poor quality he is punished and the regime of tightened inspection is entered. If he doesn't pull up his socks while in tightened inspection, the inspection scheme is discontinued and its time to call in the lawyers to cancel the whole contract.

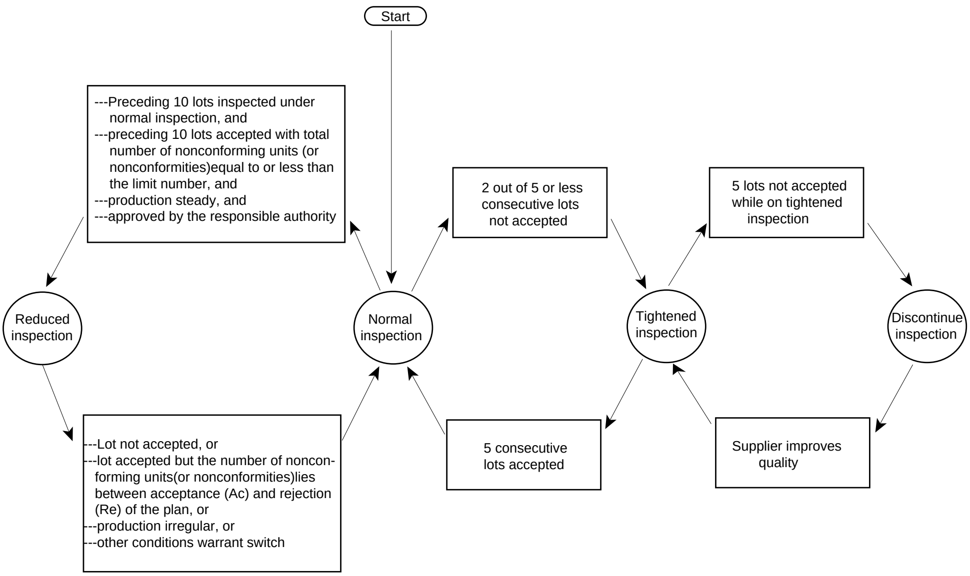

Fig. 1.4 The regimes of inspection under MIL-STD-105D.

Suppose we start with normal inspection. We read off the parameters from the Table II-A on page 10 across from code letter J. The sample size from each lot should be 80 and we accept the lot if no more than c=5 nonconforming units are in the sample. If there are 6 or more we reject the lot and send it back to the supplier. We stay with normal inspection until one of the conditions is met in Figure 1.4. Suppose the conditions for tightened inspection are met; that is two out of five or worse consecutive lots have been non-acceptable. In this case the sampling scheme given in Table II-B on page 11 applies to subsequent lots. Reading off the parameters we see the sample size is still 80 but the buyer accepts the lot if no more than c=3 nonconforming units are in the sample; otherwise the buyer rejects the lot. This puts the supplier in a tight spot. He knows that while in tight inspection he dare not have 5 more lots rejected or the sampling plan will be suspended and his contract is at risk. He will try very hard to supply good quality and return to normal inspection by having 5 lots in a row accepted.

The conditions for entering the reduced inspection regime are rather stringent. The preceding 10 lots inspected under normal inspection must be accepted. Next, the total number of nonconforming units in these 10 samples must be less than or equal the limit number in Table II-D on page 13. Hence the total number of nonconforming units in these 10 lots must be less than or equal to 14 since 10 samples of size 80 units or 800 in all were sampled. Moreover the production must be at a steady rate and finally some responsible authority must give an OK. Suppose the conditions for reduced inspection are met. In this case the sampling scheme given in Table II-C on page 12 applies to subsequent lots. Reading off the parameters we see the sample size is 32. This means less work for the buyer because he trusts the supplier. The buyer rejects the lot and returns to normal inspection if there are 5 or more nonconforming units. If no more than c=4 nonconforming units are in the sample the buyer accepts the lot but only remains in the reduced inspection regime if the number of nonconforming units is no more than 2 and production is regular and no unwarranted conditions are observed.

The process average is the percentage of nonconforming units found in the samples submitted. If the proportion of nonconforming units in a lot is $p\%$, the OC curve at $p\%$ gives the probability the lot will be accepted. Hence, if the process average is $p\%$ then in the long run a proportion $OC(p\%)$ of the lots will be accepted. The OC curves determined for the normal, tightened and reduced sampling plans are such that $OC(2.5\%) \approx 95\%$. The precise values can be calculated as we do for the scheme for testing shells in Example 3.4. This means that if the supplier maintains a process average of at most 2.5% nonconforming then 95% of his lots will be accepted.

Table II-A - Single sampling plans for normal inspection

Acceptable quality levels (normal inspection)

sample size code letter	sample size	.010	.015	.025	.040	.065	.10	.15	.25	.40	.65	1.0	1.5	2.5	4.0	6.5	10	15	25	40	65	100	150	250	400	650	1000
		Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re
A	2	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇓	⇓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31
B	3	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇑	⇓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31	44 45
C	5	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇑	⇓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31	44 45	⇑
D	8	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇑	⇓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31	44 45	⇑	⇑
E	13	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇑	⇓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31	44 45	⇑	⇑	⇑
F	20	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇑	⇓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	⇑	⇑	⇑	⇑	⇑	⇑
G	32	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇑	⇓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	⇑	⇑	⇑	⇑	⇑	⇑	⇑
H	50	⇓	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇑	⇓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑
J	80	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇑	⇓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑
K	125	⇓	⇓	⇓	⇓	⇓	0 1	⇑	⇓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑
L	200	⇓	⇓	⇓	⇓	0 1	⇑	⇓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑
M	315	⇓	⇓	⇓	0 1	⇑	⇓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑
N	500	⇓	⇓	0 1	⇑	⇓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑
P	800	⇓	0 1	⇑	⇓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑
Q	1,250	0 1	⇑	⇓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑
R	2,000	⇑	⇑	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑

⇓ = Use first sampling plan below arrow. If sample size equals or exceeds, lot or batch size, carry out 100 % inspection.

⇑ = Use first sampling plan above arrow.

Ac = Acceptance number, Re = Rejection number

Table II-B - Single sampling plans for tightened inspection

sample size code letter	sample size	Acceptable quality levels (tightened inspection)																									
		.010	.015	.025	.040	.065	.10	.15	.25	.40	.65	1.0	1.5	2.5	4.0	6.5	10	15	25	40	65	100	150	250	400	650	1000
		Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re
A	2	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	27 28
B	3	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇓	⇓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	27 28	41 42
C	5	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇓	⇓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	27 28	41 42	⇑
D	8	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇓	⇓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	27 28	41 42	⇑	⇑
E	13	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇓	⇓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	27 28	41 42	⇑	⇑	⇑
F	20	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇓	⇓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	⇑	⇑	⇑	⇑	⇑	⇑
G	32	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇓	⇓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	⇑	⇑	⇑	⇑	⇑	⇑	⇑
H	50	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇓	⇓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑
J	80	⇓	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇓	⇓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑
K	125	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇓	⇓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑
L	200	⇓	⇓	⇓	⇓	⇓	0 1	⇓	⇑	1 2	2 3	3 4	5 6	8 9	12 13	18 19	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑
M	315	⇓	⇓	⇓	⇓	0 1	⇓	⇓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑
N	500	⇓	⇓	⇓	0 1	⇓	⇓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑
P	800	⇓	⇓	0 1	⇓	⇓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑
Q	1,250	⇓	0 1	⇓	⇓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑
				⇓								⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑
R	2 ,000	0 1	⇑	⇓	1 2	2 3	3 4	5 6	8 9	12 13	18 19	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑
S	3,150			1 2																							

⇓ = Use first sampling plan below arrow. If sample size equals or exceeds, lot or batch size, carry out 100 % inspection.

⇑ = Use first sampling plan above arrow.

Ac = Acceptance number, Re = Rejection number

Table II-C - Single sampling plans for reduced inspection

Acceptable quality levels (reduced inspection)†

sample size code letter	sample size	.010	.015	.025	.040	.065	.10	.15	.25	.40	.65	1.0	1.5	2.5	4.0	6.5	10	15	25	40	65	100	150	250	400	650	1000
		Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re	Ac Re
A	2	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇓	⇓	1 2	2 3	3 4	5 6	7 8	10 11	14 15	21 22	30 31
B	2	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇑	⇓	0 2	1 3	2 4	3 5	5 6	7 8	10 11	14 15	21 22	30 31
C	2	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇑	⇓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	14 17	21 24	⇑
D	3	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇑	⇓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	14 17	21 24	⇑	⇑
E	5	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇑	⇓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	14 17	21 24	⇑	⇑	⇑
F	8	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇑	⇓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	⇑	⇑	⇑	⇑	⇑	⇑
G	13	⇓	⇓	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇑	⇓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	⇑	⇑	⇑	⇑	⇑	⇑	⇑
H	20	⇓	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇑	⇓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑
J	32	⇓	⇓	⇓	⇓	⇓	⇓	0 1	⇑	⇓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑
K	50	⇓	⇓	⇓	⇓	⇓	0 1	⇑	⇓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑
L	80	⇓	⇓	⇓	⇓	0 1	⇑	⇓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑
M	125	⇓	⇓	⇓	0 1	⇑	⇓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑
N	200	⇓	⇓	0 1	⇑	⇓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑
P	315	⇓	0 1	⇑	⇓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑
Q	500	0 1	⇑	⇓	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑	⇑
R	800	⇑ ⇑	⇑ ⇑	0 2	1 3	1 4	2 5	3 6	5 8	7 10	10 13	⇑ ⇑	⇑ ⇑	⇑ ⇑	⇑ ⇑	⇑ ⇑	⇑ ⇑	⇑ ⇑	⇑ ⇑	⇑ ⇑	⇑ ⇑	⇑ ⇑	⇑ ⇑	⇑ ⇑	⇑ ⇑	⇑ ⇑	⇑ ⇑

⇓ = Use first sampling plan below arrow. If sample size equals or exceeds, lot or batch size, carry out 100 % inspection.
⇑ = Use first sampling plan above arrow, Ac = Acceptance number, Re = Rejection number
† = If the acceptance number has been exceeded, but the rejection number has not been reached, accept the lot, but revert to normal inspection (see 11.1.4).

Table II-D - Limit numbers for reduced inspection

Acceptable quality levels (reduced inspection)†

Number of units sampled in last 10 lots	AQL 0.1	AQL 0.25	AQL 1.0	AQL 2.5	AQL 4.0	AQL 10.0
20 to 29	*	*	*	*	*	0
30 to 49	*	*	*	*	*	0
50 to 79	*	*	*	*	0	2
80 to 129	*	*	*	*	0	4
130 to 199	*	*	*	0	2	7
200 to 319	*	*	0	2	4	14
320 to 499	*	*	0	4	8	24
500 to 799	*	*	2	7	14	40
800 to 1,249	*	0	4	14	24	68
1,250 to 1,999	*	0	7	24	40	110
2,000 to 3,149	0	2	14	40	68	181
3,150 to 4,999	0	4	24	67	111	
5,000 to 7,999	2	7	40	110	181	
8,000 to 12,499	4	14	68	181		
12,500 to 19,999	4	24	110			
20,000 to 31,499	14	40	181			
31,500 to 49,999	24	67	68	181		
50,000 over	40	110				

* Denotes the number of sample units from the last ten lots is not sufficient for reduced inspection for this AQL. In this case more than 10 lots must be used for the calculation provided that the lots used are the most recent ones in sequence. Of course all were subject to normal inspection and none were rejected while on inspection.

In some cases a rejected lot is not sent back to the supplier but is instead subject to 100% inspection and all nonconforming units are replaced by conforming units. In this case the average outgoing quality - AOQ - given the process average is $p\%$, is the average quality of the outgoing product. This includes all accepted lots, plus all lots which were not initially accepted and from which all nonconforming units were replaced by conforming units. Clearly $AOQ(p\%) = OC(p\%) \cdot (p\%)$ since a proportion $(1 - OC(p\%))$ of outgoing product has no nonconforming units. If, in fact, rejected lots are repaired then the value of the AOQ at the AQL is a useful parameter when originally negociating the AQL since it represents the true proportion of nonconforming units arriving on the factory floor. Another useful parameter, when lots are repaired, is the Average Outgoing Quality Limit - $AOQL$. The $AOQL$ is the maximum of the AOQ's for all possible process averages; i.e. $AOQL = \max\{OC(p\%) \cdot (p\%) | 0 \leq p\% \leq 100\%\}$. This is the worst case scenario for measuring the true proportion of nonconforming units arriving on the factory floor.

The AOQ associated with the sample size and sampling limits in Table II-B, Table II-C and Table II-D was calculated by computer simulations. In the exercises we suggest some term projects which illustrate just how to go about analyzing quality control schemes by simulation. It is not, however, our goal to understand the precise workings of this or any other quality control procedures through simulation but rather to develop mathematical tools for the analysis of quality control schemes. Our credo is that one theorem is worth a thousand simulations!

1.4 Exercises

The following projects may be assigned at the beginning of term. Students should form teams. It is preferable that each team have at least one member with computer experience. The main mathematical work involves concepts in Chapter 5 but the simulation part can be started immediately.

Exercise 1.1 [ATM buffer management] We shall consider a simple 2×2 ATM multiplexor. Two input trunk lines carrying noisy ATM traffic enter the switch at input ports A and B and leave from ports X and Y. The cells from input port A that exit from port X are stored in buffer AX. Those from input port B that exit from port X are stored in buffer BX. The cells from input port A that exit from port Y are stored in buffer AY. Those from input port B that exit from port Y are stored in buffer BY. All four buffers have a maximum capacity of 5 cells and excess cells are lost.

Every time slot the controller at output port X performs a round robin polling of the two buffers AX and BX. The head-of-line cell is sent from one queue and then the other. If no cell is queued at the polled buffer the second buffer is immediately polled. If it is also empty then the pointer returns to the first buffer polled. We assume the arrivals at buffers AX and BX form independent Bernoulli processes

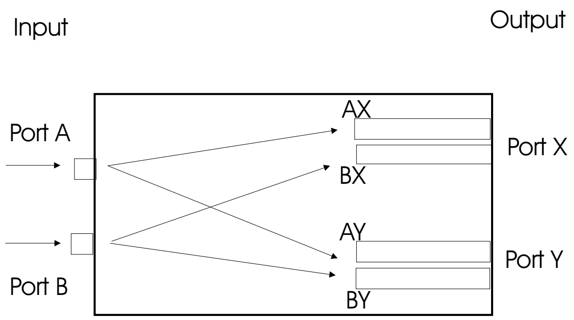

Fig. 1.5 A two by two switch

with probabilities p_A and p_B of having a cell in a given time slot.
a) Model the queue size AX and BX plus a controller pointer as a Markov chain. Write a *Mathematica* program to calculate the (72×72) transition matrix K. Calculate the cell loss probability of cells traversing output port X using K.
b) Write a computer simulation of output port X and estimate the cell loss probability. Compare with the results in a).
c) Is it possible to use another buffer management protocol other than round robin which gives a smaller cell loss probability? Make a suggestion and calculate the new cell loss probability (analytically if possible but in any case by simulation). Suggestions: Serve the Longest Queue or Serve the Oldest Cell. Discuss the disadvantages of the new protocol.
d) Discuss the difficulties of evaluating a 16×16 ATM switch.

Note that if $p_A + p_B > 1$ then on average more cells arrive than can be served. This means the buffers will tend to be overloaded and lots of cells will be lost. When $p_A + p_B$ is small however cell losses will be small. It is suggested that the simulations be done for a variety of values of p_A and p_B to get an overall picture.

Exercise 1.2 [ATM buffer management continued]
a) Find the mean time until the first cell is lost given the buffers start out empty.
b) Calculate the mean busy period which is the mean time for an empty system to become empty again.
c) Calculate the mean delay experienced by a cell which passes through buffer A.
d) Calculate the mean delay for buffer B. Use Little's law (see Chapter 6).
e) Use the simulator already developed to obtain estimates for the performance measures calculated analytically in a), b) and c). Note that the Little's law applies to cells that are queued in the system not those that are discarded so Little's law fails when the buffer is overloaded, i.e. $p_A + p_B > 1$.

Exercise 1.3 [On-line Quality Control]
a) Imagine you are to receive batches of 1000 items every week. Design a sampling acceptance scheme based on MIL-STD-105D (ISO 2859) which for an AQL of 2.5%.

b) Write a simulation for this acceptance sampling scheme. Check that the AQL of 2.5% is indeed the highest percentage of defectives acceptable as a process average. Estimate the percentage of lots rejected if this average is maintained. Next experiment to find the lot tolerance percent defective LTPD specified by this scheme. The LTPD is usually taken to be that incoming quality above which there is less than a 10% chance a lot will be accepted.
c) Suppose now that you are the manufacturer producing the above items. Suppose long experience has taught that a 1% rate of defective items is inevitable without an expensive redesign of the plan. To monitor production a item is selected at random out of every 10 items produced. Once 5 items have been selected, the items are inspected for defects. Design a Shewhart p-chart (see Example 3.27) to monitor that production is in control with a 1% rate of defectives. What is the distribution of the number of items produced before a false alarm is signaled.
d) Suppose that at some point in the future the process goes out of control and the rate of defectives increases to 5%. What is the distribution of the number of items produced after this change point before an out of control alarm is signaled.
e) Write a simulation to design a Cusum procedure (see Example 5.39) based on the lengths of runs of nondefective items inspected between occurrences of defective items. Set the on-target run length to be the same as for the p-chart.
f) Design the anchor value to minimize the mean time to signal an out of control situation if indeed the rate of defectives suddenly jumps to 5%.
g) The above Cusum procedure can be modelled by a Markov chain with forbidden out of control states. Use *Mathematica* to calculate the expected on-target and off-target run lengths of the procedure you have designed and check that these agree with simulation results.

1.5 A Probability Primer

We describe the **equiprobable** model associated with an example in data transmission. Imagine that data is generated in one, five or ten kilobit packets. On average one kilobit packets are three times more likely than ten kilobit packets and five kilobit packets are four times as likely as ten kilobit packets. Of course the transmissions might be strictly deterministic. It might be that the pattern of transmissions is always $5, 1, 5, 1, 5, 1, 5, 10$ kilobit packets repeated over and over. This deterministic model certainly describes the average flow of bytes across the network but it misses an essential component. Suppose a node in the network receives and retransmits the packets and at all times stores the last three packets. Suppose the capacity of the node is 25 kilobits. The deterministic flow poses no problem since at the worst the node must store 20 kilobits. The problem of congestion occurs when the packets arrive in random order and in this case the node might need a capacity of 30 kilobits.

Suppose we wish to describe the outcome of ten consecutive transmissions with a random arrival of packets. We do a thought experiment. Consider the experiment of drawing ten times with replacement from a bag containing three pennies (each penny represents a one kilobit packet), four nickels (each nickel represents a five kilobit packet) and a dime (representing a ten kilobit packet). Each of the eight coins is assumed to have an equal chance of being picked in any given draw. This is the model of random or probability sampling. The probabilist's job is to describe the likelihood of possible outcomes of this sampling procedure given the contents of the bag. The statistician's job is much harder since he is not told the contents of the bag and must infer its contents from the sample. In other words the probabilist usually knows the distribution of the values of the coins.

The outcome of such an experiment is random or stochastic since it can't be predicted. If we put imaginary numbers on the three pennies and the four nickels we get a list or population

$$\mathcal{L} := \{p_1, p_2, p_3, n_1, n_2, n_3, n_4, d\}$$

of possible equally likely outcomes from a single draw. We are only interested in a single aspect of each element of this population; namely the monetary value. The distribution of any aspect of a population is often represented by the mass function which is simply a function p giving the proportion $p(x)$ of the population having value x. In this case, p is given by the proportions of the number of pennies, nickels and dimes:

$$\begin{array}{r} x: 1 \;\; 5 \;\; 10 \text{ otherwise} \\ p(x): \frac{3}{8} \;\; \frac{4}{8} \;\;\; \frac{1}{8} \;\; 0 \quad\quad\;\; \end{array}$$

We often summarize this information in the population histogram as shown in Figure 1.6.

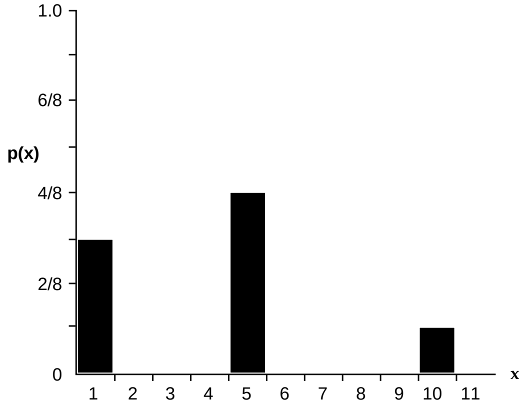

Fig. 1.6 The population histogram

The mean μ of the population is just the average value of the population which is

$$\mu = (1+1+1+5+5+5+5+10)/8 = 33/8.$$

We remark that this is just $\sum xp(x)$. The variance of the population σ^2 is the average squared deviation of the population values from the population average:

$$\sigma^2 = \frac{3(1-\mu)^2 + 4(5-\mu)^2 + (10-\mu)^2}{8}$$
$$= \sum (x-\mu)^2 p(x) = 8.359375.$$

The mean is a kind of center of the population histogram. The standard deviation σ is a measure of the spread. If, in general, the aspect values of the elements of the population are a list of numbers $(s_1, s_2, \ldots, s_N)$ (instead of $(1,1,1,5,5,5,5,10)$), we could again construct the population histogram and the corresponding population mass function p. In general,

$$\mu = \sum_{i=1}^{N} s_i/N = \sum_x xp(x) \text{ and } \sigma^2 = \sum_{i=1}^{N}(s_i-\mu)^2/N = \sum_x (x-\mu)^2 p(x).$$

Chebyshev's lemma provides a precise description of the spread of values around the mean of the list:

Lemma 1.1 *The proportion of the values in a list lying at least k standard deviations from the mean is less than $1/k^2$ for any $k > 0$.*

Proof: Those elements of the list of values lying at least k standard deviations from the mean may be written as $F := (s : |s-\mu| \geq k \cdot \sigma)$. The proportion of elements in the list in F is $\sharp(F)/N$ where $\sharp(F)$ denotes the number of elements in F. Now by definition

$$\begin{aligned}\sigma^2 &= \frac{\sum_{i=1}^{N}(s_i-\mu)^2}{N} \geq \frac{\sum_{s_i\in F}(s_i-\mu)^2}{N}\\ &\geq \frac{\sum_{s_i\in F}(k\sigma)^2}{N}\\ &= k^2\sigma^2\frac{\sharp(F)}{N}.\end{aligned}$$

Now divide through by $k^2\sigma^2$ and we have $\sharp(F)/N \leq 1/k^2$ as desired. ■

Applying Chebyshev's lemma to the above population we get that the proportion of the population lying at least 2 standard deviations from the mean $\mu = 33/8 = 4.125$ is less than $1/4$. Since $\sigma = 2.8912$ approximately, we see $\mu + 2\sigma = 9.9075$ hence only the value 10 lies this far from the mean. The element 10 represents only $1/8$ of the list so the inequality is verified.

Now suppose we wish to describe 10 draws from $\mathcal{L}$. A list of all possible outcomes is called the sample space and is described by

$$\Omega = \{\omega = (x_1, x_2, \ldots, x_{10}) : x_i \in \mathcal{L}; i = 1, 2, \ldots, 10\}$$

where ω is used to denote a typical outcome or sample point and p_1 denotes the first penny, p_2 the second and so on. One particular sample point might be

$$\omega_0 = (p_1, d, p_2, n_2, d, p_1, p_3, n_4, n_2, n_3).$$

This corresponds to first drawing penny number 1 then the dime and then penny number 2 and so on up to nickel number 3 on tenth draw. Later we will call Ω a product space and we will use the notation

$$\Omega = \{p_1, p_2, p_3, n_1, n_2, n_3, n_4, d\}^{10} = \mathcal{L}^{10}.$$

By symmetry each outcome is equally likely. The number of points in Ω is $\sharp(\Omega) = 8^{10}$ since each of the 10 draws could be one element of the 8 element set $\mathcal{L}$. Hence, intuitively, the probability of each sample point is 8^{-10}. The probability of an event of interest like the probability of getting at least one dime is clearly proportional to the number of sample points in the event. Let A represent the event or subset of Ω corresponding to getting at least one dime. The sample point ω_0, for instance, is in A. Let $P(A)$ denote the probability that the outcome of the random experiment falls in the event A; so

$$P(A) = \frac{\sharp(A)}{\sharp(\Omega)}.$$

For equiprobable models such as this one, calculating probabilities reduces to counting. This is the reason we chose the sample space Ω above. If, instead, we had chosen $\Omega = \{p, n, d\}^{10}$ we would not have an equiprobable model and the construction of the appropriate P to describe random drawing requires more care (see the product probability in Chapter 2).

Counting is not so easy! We shall assume everybody knows permutations and combinations but it is not immediately clear how to easily calculate the probability of getting at least one dime. It is useful to develop a series of axioms about probabilities to make things easier. This will be done systematically in Chapter 2.

First, $P(\Omega) = 1$. Next, if two events A and B are disjoint, that is they have no elements in common, then the probability of their union is the sum of their probabilities, i.e. $P(A \cup B) = P(A) + P(B)$. This follows since

$$P(A \cup B) = \frac{\sharp(A \cup B)}{\sharp(\Omega)} = \frac{\sharp(A) + \sharp(B)}{\sharp(\Omega)} = \frac{\sharp(A)}{\sharp(\Omega)} + \frac{\sharp(B)}{\sharp(\Omega)}$$
$$= P(A) + P(B).$$

The complement of A, i.e. those points which are not in A, is denoted by A'. A' has probability $P(A') = 1 - P(A)$. This follows immediately from the fact that $1 = P(\Omega) = P(A + A') = P(A) + P(A')$ since A and A' are disjoint. Consequently, if A' represents the event where no dimes are drawn then $P(A) = 1 - (7/8)^{10}$ since the number of sample points in A' is 7^{10} (each draw can be chosen from $\{p_1, p_2, p_3, n_1, n_2, n_3, n_4\}$).

The intuitive use of conditional probabilities is what separates probabilists from measure theorists. If, in our example, we know that the event B of drawing exactly 3 dimes has occurred (or exactly 3 ten kilobit packets have arrived among the last 10 packets), what is the probability of the event C that we draw 3 dimes in a row (or equivalently, that the node is overloaded by 3 ten kilobit packets in a row)? By symmetry, the probability that an outcome in C occurs given that B has occurred is the proportion of the number of sample points in both B and C divided by the number in B; that is we define the conditional probability of C given B to be

$$P(C|B) = \frac{\sharp(B \cap C)}{\sharp(B)} = \frac{\sharp(B \cap C)/\sharp(\Omega)}{\sharp(B)/\sharp(\Omega)} = \frac{P(B \cap C)}{P(B)}.$$

If we work this out we see $\sharp(B)$ is the number of ways of choosing exactly 3 dimes from the 10 draws times the number of ways of drawing each of the remaining 7 coins from $\{p_1, p_2, p_3, n_1, n_2, n_3, n_4\}$. Hence,

$$\sharp(B) = \binom{10}{3} \times 7^7 \text{ so } P(B) = \binom{10}{3}\left(\frac{1}{8}\right)^3\left(\frac{7}{8}\right)^7.$$

Similarly $\sharp(B \cap C)$ is the number of ways of choosing 3 consecutive draws to be dimes; that is 8 ways times 7^7, the number of ways of drawing each of the remaining 7 coins from $\{p_1, p_2, p_3, n_1, n_2, n_3, n_4\}$. Hence, $\sharp(B \cap C) = 8 \times 7^7$. We conclude

$$P(C|B) = 8/\binom{10}{3} = \frac{1}{15}.$$

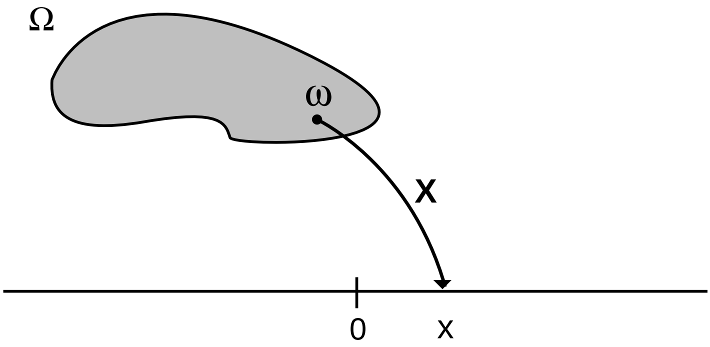

Fig. 1.7 A random variable

The idea of independence is obvious. We say an event F is independent of an event E if the probability that F will occur is independent of knowing whether E occurred or not; that is $P(F|E) = P(F)$. Multiplying by $P(E)$ and using the definition $P(F|E) := P(E \cap F)/P(E)$, we see this is equivalent to $P(E \cap F) = P(E) \cdot P(F)$. In particular if E is the event that the first coin is a nickel and F is the event that the second coin is a dime then clearly we have independent events and this can be checked by calculation.

1.6 A Random Sample

Let $X_1, \ldots, X_{10}$ represent the values of the coins drawn (or packets transmitted) during the first, through tenth draws. These *random variables* are defined on the sample space Ω. For the sample point $\omega_0 = (p_1, d, p_2, n_2, d, p_1, p_3, n_4, n_2, n_3)$, $X_1(\omega_0) = 1, X_2(\omega_0) = 10, X_3(\omega_0) = 1$ and so on. In general a random variable calculates some aspect of a sample point. For instance we might define X to be the total number of dimes drawn so $X(\omega_0) = 2$. Figure 1.7 illustrates a random variable X defined at each point ω of a sample space.

The description of a random variable starts with the range, $\mathcal{R}_X$, of values taken on. For instance $\mathcal{R}_X = \{0, 1, 2, \ldots, 10\}$. Next we specify the likelihood of these values by constructing the histogram of the list of values $X(\omega)$ for $\omega \in \Omega$. Since this list has 8^{10} elements this might seem difficult but when we group ω's giving the same X value we see the histogram is equivalent to the probability mass function or p.m.f.

$$p_X(x) := P(\{\omega : X(\omega) = x\}) \equiv P(X = x), \text{ for } x \in \mathcal{R}_X.$$

For instance, $p_X(3)$ is the probability precisely three dimes are drawn and this has

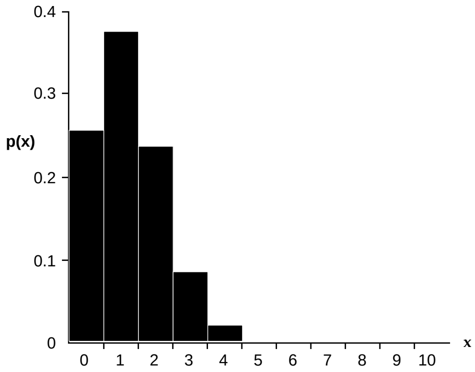

Fig. 1.8 The histogram of X

been calculated above. By the same reasoning we can get

$$p_X(x) = \binom{10}{x} (\frac{1}{8})^x (\frac{7}{8})^{10-x} \text{ for } x \in \{0, 1, \ldots, 10\}.$$

This is a binomial distribution investigated more thoroughly later. The histogram of the random variable X is given in Figure 1.8.

The p.m.f of X_1 is obviously

$$\begin{array}{rl} x : & 1 \;\; 5 \;\; 10 \text{ otherwise} \\ p_{X_1}(x) : & \frac{3}{8} \; \frac{4}{8} \;\; \frac{1}{8} \;\; 0. \end{array}$$

Formally we may get these probabilities by counting. Recall that we denote an arbitrary point $(x_1, x_2, \ldots, x_{10}) \in \Omega$ by ω. Hence

$$\begin{aligned} P(X_1 = 5) &= P(\{\omega : x_1 \in \{n_1, n_2, n_3, n_4\}, x_i \in \mathcal{L}; i = 2, \ldots, 10\}) \\ &= \sharp(\{\omega : x_1 \in \{n_1, n_2, n_3, n_4\}, x_i \in \mathcal{L}; i = 2, \ldots, 10\})/\sharp(\Omega) \\ &= 4 \times 8 \times 8 \times \cdots \times 8/8^{10} \\ &= 4/8. \end{aligned}$$

By similar reasoning we see the p.m.f.'s of $X_1, X_2, \ldots, X_{10}$ are all the same and are *equal* to the p.m.f. of the population p. This is the link between the real and perhaps unknown (to the statistician) distribution of the population and the sample which we observe. Each of the sampled values represents the population in the sense that its distribution or p.m.f. is that of the population.

The sequence of random variables $X_1, X_2, \ldots, X_{10}$ is an *i.i.d.* sequence; that is the random variables are *i*ndependent and *i*dentically *d*istributed where we say random variables are independent if events generated by the individual X_i's are independent in the sense given above. To be more precise we let $\sigma(X_i)$ be the collection of events in Ω of the form $\{X_i \in H_i\}$ where H_i is some subset of real numbers $\mathcal{R}$. For instance take $H_1 = \{5\}$. Then the event $\{X_1 \in H_1\} \equiv \{\omega :$

$X_1(\omega) = 5\}$ is the set investigated above. The event of drawing a nickel on the first draw is therefore in $\sigma(X_1)$. Similarly, taking $H_1 = \{1, 10\}$, we see that the event of not drawing a nickel on the first draw is also in $\sigma(X_1)$. If we let $H_2 = \{10\}$ then

$$\begin{aligned}\{X_2 \in H_2\} &= \{\omega = (x_1, x_2, \ldots, x_{10}) : X_2(\omega) = 10\} \\ &= \{\omega : x_2 = d, x_i \in \mathcal{L}; i = 1, 3, 4, \ldots, 10\}\end{aligned}$$

is in $\sigma(X_2)$. Clearly $P(X_2 \in H_2) = 1/8$. Moreover

$$\begin{aligned}&P(\{X_1 \in H_1\} \cap \{X_2 \in H_2\}) \\ &= P(\{\omega : x_1 \in \{n_1, n_2, n_3, n_4\}, x_2 = d, x_i \in \mathcal{L}; i \geq 3\}) \\ &= \frac{4 \times 1 \times 8 \times \cdots \times 8}{8 \times 8 \times 8 \times \cdots \times 8} \\ &= \frac{4}{8} \cdot \frac{1}{8} \\ &= P(X_1 \in H_1) \cdot P(X_2 \in H_2).\end{aligned}$$

This is the formal proof that the event of drawing a nickel on the first draw and a dime on the second are independent. It is clear that any events of the form $\{X_1 \in H_1\}$ and $\{X_2 \in H_2\}$ are independent by the same reasoning and we therefore declare the random variables X_1 and X_2 to be independent. By the same token we see all the X_i's are mutually independent.

We shall need to express the collection of events that are determined by observing a sequence of random variables. Define $\mathcal{F}_m := \sigma(X_1, X_2, \ldots, X_m)$ to be the set of events of the form

$$E = \{\omega : (X_1(\omega), X_2(\omega), \ldots, X_m(\omega)) \in H\}$$

where H is a subset of $\Re^m$. This just means we can determine if an ω is in E or not by observing the values of $X_1(\omega), \ldots, X_m(\omega)$. We call $\mathcal{F}_m$ the past of the random sequence up to observation or time m.

The expected value of a random variable is the average of the list of its values:

$$EX \equiv \mu_X := \frac{\sum_{\omega \in \Omega} X(\omega)}{\sharp(\Omega)} = \sum_{x \in \mathcal{R}_X} x p_X(x).$$

The latter equality is obtained by grouping together all those ω which are sent to a common value x; i.e. $\{\omega : X(\omega) = x\}$. Clearly the contribution of these points to the average is just $x \cdot \sharp(\{\omega : X(\omega) = x\})$. However $p_X(x) = P(X = x) = \sharp(\{\omega : X(\omega) = x\})/\sharp(\Omega)$ which gives the result. For instance, if X represents the number of dimes drawn then

$$EX = \sum_{x=0}^{10} x \binom{10}{x} (\frac{1}{8})^x (\frac{7}{8})^{10-x}.$$

This result had better be 10/8 (and it is) since there is one chance in eight of drawing a dime and we draw ten times. The expected value of X_1 is clearly equal to $\mu = 33/8$ since the histogram of X_1 is the same as that of the population.

The expected value is useful because it measures the center of the histogram of X. We can also calculate the expected value of the sum $X+Y$ where Y is another random variable. By definition

$$E(X+Y) = \frac{\sum_{\omega\in\Omega}(X(\omega)+Y(\omega))}{\sharp(\Omega)} = \frac{\sum_{\omega\in\Omega}X(\omega)}{\sharp(\Omega)} + \frac{\sum_{\omega\in\Omega}Y(\omega)}{\sharp(\Omega)}$$
$$= EX + EY.$$

Also, if c is a constant value

$$E(cX) = \frac{\sum_{\omega\in\Omega}cX(\omega)}{\sharp(\Omega)} = c\frac{\sum_{\omega\in\Omega}X(\omega)}{\sharp(\Omega)} = c \cdot EX.$$

This linearity is a very useful property of the expectation. The 50^{th} percentile (the median) also measures the center of the histogram of X but does not have this linearity property so it is much less useful. If T represents the total value of the ten draws then $ET = E(X_1 + \cdots + X_{10}) = EX_1 + \cdots + EX_{10} = 10 \cdot \mu$ again by linearity. Also if $\overline{X} := (X_1 + \cdots + X_{10})/10$ then again by linearity $E\overline{X} = \mu$. This means the histogram of the random variable $\overline{X}$ is centered at the population mean μ and $\overline{X}$ is what we call an *unbiased estimator* of μ.

We can make new random variables from old by defining functions of X like $h(x) = x^2$ or $h(x) = \max(x^2, 5)$. The expectation is still given from the definition:

$$Eh(X) = \frac{\sum_{\omega\in\Omega}h(X(\omega))}{\sharp(\Omega)} = \sum_{x\in\mathcal{R}_X} h(x)p_X(x)$$

by again grouping those ω which are sent to the same value x. This expression is called the *law of the unconscious statistician*. For instance, if X is the number of dimes drawn, then

$$E\max(X^2,5) = \sum_{x=0}^{2} 5\binom{10}{x}(\frac{1}{8})^x(\frac{7}{8})^{10-x} + \sum_{x=3}^{10} x^2\binom{10}{x}(\frac{1}{8})^x(\frac{7}{8})^{10-x}.$$

The most important application of the law of the unconscious statistician is the definition of the variance of a random variable: $\sigma_X^2 := E(X-\mu_X)^2$. Since the variance of X is precisely the variance of the list $X(\omega)$ we have, by Chebyshev's lemma, that the proportion of this list at least k standard deviations σ_X from the mean of the list μ_X is less than $1/k^2$. However the proportion of the list at least k standard deviations from the mean is precisely $P(|X-\mu_x| \geq k \cdot \sigma_X)$ so we have Chebyshev's lemma for random variables:

Lemma 1.2 *For any random variable with expected value μ_X and standard deviation σ_X we have $P(|X-\mu_x| \geq k \cdot \sigma_X) \leq 1/k^2$.*

1.7 Joint Distributions

The joint behavior of random variables and vectors is discussed in detail in Chapter 2 but here let us focus on the simple case of two variables. To be concrete let X be the number of dimes drawn and let T be the total value of the coins drawn. The joint p.m.f. is defined to be

$$p_{X,T}(x,t) = P(X = x \text{ and } T = t) = P(\{\omega : X(\omega) = x, T(\omega) = t\})$$

where $x \in \mathcal{R}_X$ and $t \in \mathcal{R}_T$. With this we can repeat most of the calculations done for one variable. In particular, if we create a new random variable such as $h(X,T)$ where h is a function like $h(x,y) = x + 2y$ then we again have a law of the unconscious statistician:

$$Eh(X,T) = \sum_{x\in\mathcal{R}_X, t\in\mathcal{R}_T} h(x,t)p_{X,T}(x,t).$$

This is proved just as before since the expected value is by definition

$$\sum_{\omega\in\Omega} h(X(\omega), Y(\omega))/\sharp(\Omega).$$

Calculating the joint p.m.f. could prove to be a lot of work. There are a few short cuts. Define the conditional p.m.f. of T given X to be

$$\begin{aligned} p_{T|X}(t|x) &:= P(T = t|X = x) = P(T = t, X = x)/P(X = x) \\ &= p_{X,T}(x,t)/p_X(x). \end{aligned}$$

If we use the law of the unconscious statistician given above we see

$$\begin{aligned} Eh(X,T) &= \sum_{x\in\mathcal{R}_X, t\in\mathcal{R}_T} h(x,t)p_{X,T}(x,t) \\ &= \sum_{x\in\mathcal{R}_X} [\sum_{t\in\mathcal{R}_T} h(x,t)p_{T|X}(t|x)]p_X(x) \\ &\equiv \sum_{x\in\mathcal{R}_X} E(h(x,T)|X = x)p_X(x) \end{aligned}$$

where $E(h(x,T)|X = x)$ denotes the expectation of the random variable $h(x,T)$ relative to the conditional probability $P(\cdot|X = x)$; that is the probability given by $P(A|X = x) = P(A \cap \{X = x\})/P(X = x)$. Sometimes we can apply our intuition to discover $p_{T|X}(t|x)$ or perhaps $E(T|X = x)$. For instance, if the number of dimes drawn is $x = 3$ then we know for sure $T > 30$. In fact the conditional distribution of T is the same as drawing 7 times from a similar sack without a dime and adding 30; that is $30 + \sum_{j=1}^{7} Y_j$ where Y_j denotes the value of the j^{th} draw with replacement from a sack containing $\{p_1, p_2, p_3, n_1, n_2, n_3, n_4\}$. Clearly $EY_j = 23/7$ so $E(T|X = 3) = 30 + 7 \cdot 23/7$.

In the case of independent variables everything simplifies. If the variables X and Y are independent, then the events $\{X = x\}$ and $\{Y = y\}$ are independent for any choice of x, y. Consequently

$$\begin{aligned} p_{X,Y}(x,y) &= P(\{X = x\} \cap \{Y = y\}) \\ &= P(X = x)P(Y = y) \\ &= p_X(x) \cdot p_Y(y); \end{aligned}$$

that is the joint p.m.f. is a product of the one dimensional or marginal p.m.f.'s.

A major bonus is the fact that the expectation of a product of independent variables is the product of the expectations:

$$\begin{aligned} E[X \cdot Y] &= \sum_{x,y} x \cdot y \; p_{X,Y}(x,y) = \sum_{x,y} x \cdot y \; p_X(x) p_Y(y) \\ &= \sum_x x p_X(x) \sum_y y p_Y(y) \\ &= EX \cdot EY. \end{aligned}$$

As a corollary, we easily see that the variance of a sum of independent variables is the sum of the variances. This is discussed in Chapter 2 and it bears repeating since this is the essential reason for defining the variance in the first place. In fact, it follows that the variance of $T = X_1 + \cdots + X_{10}$ is the sum of the variances of the components, therefore $\sigma_T^2 = 10 \cdot \sigma^2$. Since $\overline{X} = T/10$ and since

$$\sigma_{T/10}^2 := E(T/10 - \mu_T/10)^2 = \sigma_T^2/10^2 = \sigma^2/10$$

it follows that $\sigma_{\overline{X}} = \sigma/\sqrt{10}$. Now apply Chebyshev's Lemma to $\overline{X}$ with $k = \epsilon/\sigma_{\overline{X}}$ and remember $E\overline{X} = \mu$ to get

$$\begin{aligned} P(|\overline{X} - \mu| \geq \epsilon) &= P(|\overline{X} - \mu| \geq (\frac{\epsilon\sqrt{10}}{\sigma} \cdot \sigma/\sqrt{10})) \\ &\leq \frac{\sigma^2}{10\epsilon^2} \end{aligned}$$

where ϵ is any small number.

The result for a sample of size n instead of 10 is

$$P(|\overline{X} - \mu| \geq \epsilon) \leq \sigma^2/(n\epsilon^2).$$

The statistician who doesn't know the contents of the bag uses $\overline{X}$ as a guess for the mean value of the coins in the bag. We have shown this is an unbiased estimator, but the above shows that if the sample size is large enough, $\overline{X}$ is probably very close to the population mean μ. In fact as $n \to \infty$ the probability that the guess misses by more than ϵ tends to 0 no matter how small ϵ is! Moreover, the statistician can guess the population histogram from the histogram of the sample. The sample histogram is the histogram produced from the list $X_1(\omega), X_2(\omega), \ldots, X_n(\omega)$. It is equivalent to the sample p.m.f. given by $p_n(x) := \sharp(i : X_i(\omega) = x)/n$; that is

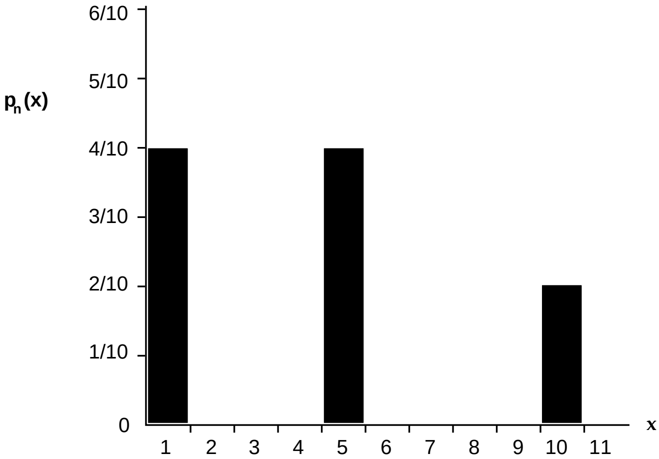

Fig. 1.9 The sample histogram

the proportion of the sample which takes on the value x. The sample histogram associated with the realization ω_0 is given in Figure 1.9.

It is reasonable to guess that the sample proportions should be close to the population proportions. If we define the random variables e_i to be 1 if $X_i = x$ and 0 otherwise, we have $p_{e_i}(1) = P(e_i = 1) = p(x)$ and $p_{e_i}(0) = P(e_i = 0) = 1 - p(x)$. Moreover $p_n(x) = \sum_{i=1}^{n} e_i/n$ so $Ep_n(x) = (Ee_1 + \cdots + Ee_n)/n = Ee_1$. However

$$Ee_1 = 1 \cdot P(X_i = x) + 0 \cdot P(X_i \neq x) = p(x)$$

so $p_n(x)$ is an unbiased estimator of the population proportion. Also the variance of $p_n(x)$, by the argument used above for $\overline{X}$, is $\sigma_{e_1}^2/n$. However $\sigma_{e_1}^2 = (1-p(x))^2 p(x) + (0 - p(x))^2(1 - p(x)) = p(x)(1 - p(x))$ so we conclude the variance of $p_n(x)$ is $p(x)(1 - p(x))/n$ which tends to 0 as $n \to \infty$. Using Chebyshev's lemma as above, we conclude that for n large enough, the probability the sample proportion $p_n(x)$ differs from the population proportion $p(x)$ by more than ϵ is vanishingly small.

This is the basis of statistical inference! Since the histogram of the sample is close to the population histogram for large n, any population parameter may be estimated by the corresponding sample parameter. The population mean μ may be estimated by the sample mean $\overline{X}$; the percentiles of the population histogram may be estimated by the sample percentiles and so on. We shall assume throughout this text that the hard work of estimating the parameters of our models has been done by somebody else, but in the real world that somebody might be you!

Even beyond the question of estimating parameters is the larger question; is the model appropriate? The arrival of packets at a node is probably not a deterministic string nicely planned to be in repeat patterns of $5, 1, 5, 1, 5, 1, 5, 10$ kilobit packets. Neither are the arrivals likely to be as random as drawing with replacement from a sack. The truth will lie somewhere in between. At best we can calculate the level of congestion associated with different models of the arrival process and make a reasonable compromise in designing the system.

1.8 Exercises

Exercise 1.1 There are 6 horses in a race. What is the probability we can correctly predict the horses that win, place and show if we pick 3 of the six horses at random?

Exercise 1.2 Let E, F, and G be three events. Express the following events in symbolic notation.
a) At least one event occurs.
b) None of these events occur.
c) At most one of these events occurs.
d) G occurs but not E or F.
e) All three events occur.
f) At most two of these events occur.

Exercise 1.3 For the counting or equiprobable model introduced in this chapter show that $P(E \cup F) \leq P(E) + P(F)$ for all events E and F.

Exercise 1.4 What is the chance of dealing a poker hand with four kings?

Exercise 1.5 Your probability class has n students. What is the probability that two or more students have the same birthday.

Exercise 1.6 Genetic theory says that the sex of a child is male or female with equal probability. We take a random sample of 100 families with two children.
a) Construct a sample space which will describe the possible outcomes.
b) How many points are in this sample space?
c) Let $X_i; i = 1, \ldots, 100$ be random variables which denote the number of girls in each of the sample families. Sketch a likely sample histogram.
d) What is the approximate mean and standard deviation of this sample histogram?

Exercise 1.7 A student is writing a multiple choice exam containing 6 questions. Each question has 6 possible responses, exactly one of which is correct. The student has spent the semester partying, and has no idea what the correct answers are. He selects answers at random from the 6 alternatives. What is the probability that he will pass the test (i.e. give 3 or more correct answers)?

Exercise 1.8 Suppose ten percent of Ford Escorts have defective head gaskets. What is the approximate probability that a dealer who buys 7 Escorts has no defective head gaskets among the 7?

Exercise 1.9 A production line produces bearings. Each bearing has a probability of 0.13 of being defective. We shall assume defects occur independently among the bearings.

a) A lot contains 13 bearings. Find the probability that this lot contains more than 2 defective bearings.
b) The production line has been in production since 8 a.m. If a bearing is produced each minute, how long would one expect to wait until the first defective bearing is produced?

Exercise 1.10 Below are the descriptive statistics of the weights of 2000 sacks of potatoes selected at random by the quality control department. The associated histogram has two bumps because there are two filling machines filling these sacks to a nominal weight of 5 kilograms. One underfills the sacks and one overfills the sacks so the histogram is really the superposition of two histograms and hence has two bumps. Shipments are made in lots of 100 sacks on a skid (you can assume the fill weights are independent).

```
Variable  N     Mean    Median   TrMean   StDev  SE Mean
C1      2000  4.9961  4.9916  4.9969   0.5496   0.0123

Variable  Minimum  Maximum         Q1          Q3
C1         3.6911   6.1061     4.4998      5.4880
```

Answer the questions in brackets below:
a) If a buyer takes a lot at random and makes a histogram of the weights of the individual sacks then this histogram will follow the normal curve (yes or no).
b) The average of this sample will be approximately (number).
c) The standard deviation of this sample histogram will be (number).
The buyer is really concerned about the total weight on a skid. Suppose he takes 75 skids at random and makes a histogram of the total weight of the sacks on each of the 75 skids then
d) this histogram will necessarily follow a normal curve (yes or no).
e) The expected value of this sample histogram will be approximately (number).
f) The standard deviation of this sample histogram will be (number).
g) What proportion of skids have a total net weight greater than 400 kilograms (number)?
h) What is the 10^{th} percentile of the total net weight on the skids (number)?

Exercise 1.11 Your company buys 70% of its light bulbs from supplier A, 20% from supplier B and 10% from supplier C. Past data has shown that 5% of the bulbs supplied by A are defective, that 3% of those supplied by B are defective and that 20% of those supplied by C are defective (company C belongs to the owner's brother-in-law).
a) A light bulb is chosen at random from your stock. What is the probability that the bulb is defective?
b) Given that the bulb chosen was in fact defective, find the probability that it

came from your brother-in-law's company.

Exercise 1.12 Past experience has shown that the parts of supplier A are just as likely to be defective as those of supplier B but those of supplier C are 3 times more likely to be defective. Purchasing records show that we have bought 20% of our parts from A, 50% from B and 30% from C. An article has been returned because of a faulty part. What is the probability that the part came from supplier A?

Exercise 1.13 If we take a sample of size $1,000$ from a population with mean 1 and variance 16, what is the probability $|\overline{X} - 1|$ is greater than 0.5?

Exercise 1.14 Suppose 30 packets are stored at a DataPac node. These packets are randomly distributed according to the distribution of Figure 1.6. Give an upper bound on the probability that more than 235 kilobytes of storage will be required.