8. Limit results for sequences of random variables

I know of scarcely anything so apt to imagination
as the wonderful form of cosmic order expressed by the law of errors.
The huger the mob and the greater the anarchy, the more perfect is its way.
Francis Galton

Here we will consider some ideas about the long run behavior of random variables and their distributions.

8.1. Convergence of random variables

Let X be a random variable and let X_1, X_2, ... be a sequence of random variables. Since these are really just functions over Ω, we can simply apply the definition of convergence of a function from analysis. The sequence X_n converges to X (written $X_n \to X$) if

$$\lim_{n\to\infty} X_n(\omega) = X(\omega) \text{ for every } \omega \in \Omega.$$

In other words, $X_n \to X$ if for every $\omega \in \Omega$ and $\varepsilon > 0$, there exists an N such that $|X_n(\omega) - X(\omega)| < \varepsilon$ for all $n \geq N$. A more intuitive way to view the definition is that no matter what happens in the real world (you can think of a random variable X as a “black box”: you tell it what happened in the real world ($\omega \in \Omega$), and it will give you a number back $X\colon \Omega \to R^1$), the random variables take on values such that the sequence converges. In probability theory, this is called **sure convergence**.

If we weaken the requirements slightly, we arrive at a second notion of convergence. The sequence X_n **converges almost surely** (or **converges with probability one**) to X (written $X_n \xrightarrow{\text{a.s.}} X$) if

$$P(\omega \in \Omega\colon X_n(\omega) \to X(\omega)) = 1,$$

or alternatively, if

$$P(\omega \in \Omega\colon X_n(\omega) \nrightarrow X(\omega)) = 0.$$

In other words, $X_n \xrightarrow{\text{a.s.}} X$ if the set of events that do not lead to convergence has measure zero. Almost sure convergence is used much more often in probability theory than sure convergence.

There is a still weaker form of convergence which relies even more on probabilistic ideas. The sequence X_n **converges in probability** to X (written $X_n \xrightarrow{P} X$) if for every $\varepsilon > 0$

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0.$$

In other words, $X_n \xrightarrow{P} X$ if for every $\delta > 0$ and $\varepsilon > 0$, there exists an N such that $P(|X_n - X| > \varepsilon) < \delta$ for all $n > N$.

Convergence in probability essentially means that the probability that $|X_n - X|$ exceeds any prescribed, strictly positive value converges to zero. The basic idea behind this type of convergence is that the probability of an "unusual" outcome becomes smaller and smaller as the sequence progresses. A sequence of random variables that converges in probability can still have an infinite number of violations of the convergence inequality.

Almost sure convergence implies convergence in probability, although the latter is used more often in introductory textbooks because it is usually easier to demonstrate for a given sequence.

The concept of convergence in probability is used very often in statistics. For example, an estimator is called **consistent** if it converges in probability to the parameter being estimated.

The final mode of convergence relies exclusively on the idea of probability. Let $F_n(a) = P(X_n \le a)$ and $F(a) = P(X \le a)$.

The sequence X_n **converges in distribution** to X (written $X_n \xrightarrow{D} X$) if

$$\lim_{n\to\infty} F_n(a) = F(a).$$

for all a such that $F(a)$ is continuous. Convergence in distribution (also called **weak convergence**) forms the basis of the central limit theorem.

Convergence in probability implies convergence in distribution.

Thus, we have a "hierarchy" of convergence definitions:

$$X_n \xrightarrow{\text{a.s.}} X \;\Rightarrow\; X_n \xrightarrow{P} X \;\Rightarrow\; X_n \xrightarrow{D} X.$$

For convergence in distribution, it makes no difference whether the random variables X_n are independent or not; they do not even need to be defined on the same probability space. On the other hand, almost sure convergence implies a strong form of dependence between the random variables involved.

8.2. Inequalities

Calculate the exact probability that X lies in some set of interest is not always easy. However, simple bounds on these probabilities will often be sufficient for the task in hand.

We start with a basic inequality.

Theorem 8.1 (Basic inequality). *If $h(\cdot)$ is a nonnegative function, then, for any $a > 0$*

$$P(h(X) \ge a) \le E(h(X))/a. \tag{8.1}$$

Proof. Define the following function:

$$I_{(h \ge a)}(\omega) = \begin{cases} 1, & h(X) \ge a, \\ 0, & \text{otherwise.} \end{cases}$$

I_A is an **indicator** of a set A, and $E(I) = P(h(X) \ge a)$. Now, by its construction the indicator I satisfies $h(X) - aI \ge 0$, and $E(h(X)) \ge aE(I) = aP(h(X) \ge a)$.

The following useful inequalities can all be proved using Theorem 8.1 or by essentially the same method. You should do them as exercises. For any $a > 0$, we have:

Markov's inequality

$$P(|X| \geq a) \leq E(|X|)/a; \tag{8.2}$$

Chebyshov's inequalitiy†

$$P(|X| \geq a) \leq E(X^2)/a^2. \tag{8.3}$$

Chebyshov's inequality is perhaps the most famous in the whole probability theory (and probably the most famous achievement of the prominent Russian mathematician P.L. Chebyshov (1821–1894)).

It follows from (8.3)

$$P(|X - E(X)| \geq a) \leq \text{var}\,(X)/a^2. \tag{8.4}$$

The domain of applications of these inequalities is huge (and not restricted to probability theory).

The names of P.L. Chebyshov and A.A. Markov (Chebyshov's pupil A.A. Markov (1856–1922) is another prominent Russian mathematician) are associated with the rise of the Russian (more precisely, St. Petersburg) school of probability theory. Neither of them could be described as having an ordinary personality. P.L. Chebyshov had wide interests in various branches of contemporary science (and also in the political, economical and social life of the period). This included the study of ballistics in response to demands by his brother who was distinguished artillery general in the Russian Imperial Army. A.A Markov was a well-known liberal opposed to the tsarist regime: in 1913, when Russia celebrated the 300th anniversary of the Imperial House of Romanov, he and some of his colleagues defiantly organized a celebration of the 200th anniversary of the law of large numbers.

Here is one important application.

Example. Let X be a random variable such that var $(X) = 0$. Show that X is constant with probability one, i.e., $P(X = E(X)) = 1$.

Solution. By (8.4), for any integer $n \geq 1$,

$$P[|X - E(X)| > 1/n] \leq n^2 \text{var}\,(X) = 0. \tag{8.5}$$

Hence, defining the events $C_n = \{|X - E(X)| > 1/n\}\uparrow$, we have (see Appendix 5)

$$P(X \neq E(X)) = P(\cup_{n=1}^{\infty} C_n) = P\left(\lim_{n\to\infty} C_n\right) = \lim_{n\to\infty} P(C_n) = 0.$$

Another example of a powerful inequality used in more than one area of mathematics is **Jensen's inequality**. It is named after J.L. Jensen (1859–1925), a Danish analyst who used it in his 1906 year paper.

It is connected with an important concept of *convexity* that crops up in many areas of pure and applied mathematics.

Definition. *A function $g(\cdot)$ (from R^1 to R^1) is called* **convex** *if, for all a, there exists $\lambda(a)$ such that*

$$g(x) \geq g(a) + \lambda(a)(x - a), \text{ for all } x. \tag{8.6}$$

If $g(\cdot)$ is differentiable, then a suitable λ is given by $\lambda(a) = g'(a)$ and (8.6) takes the form

$$g(x) \geq g(a) + g'(a)(x - a). \tag{8.7}$$

This says that a convex function lies above all its tangents. If $g(\cdot)$ is not differentiable, then there may be many choices for λ; draw a picture of $g(x) = |x|$ at $x = 0$ to see this. (There are several other definitions of a convex function, all equivalent to this.) We are interested in the following property of convex functions.

Theorem 8.2 (Jensen's inequality). *Let X be a random variable with finite mean and $g(\cdot)$ be a convex function. Then,*

$$E(g(X)) \geq g(E(X)). \tag{8.8}$$

Proof. Choosing $a = E(X)$ in (8.6), we have $g(X) \geq g(E(X)) + \lambda(X - E(X))$.

Taking the expected value of each side gives (8.8).

For example, $g(x) = |x|$ and $g(x) = x^2$ are both convex, so $E(|X|) \geq |E(X)|$ and $E(X^2) \geq (E(X))^2$, $E(\log X) \leq \log E(X)$ (because $-\log x$ is convex).

The inequality has many important applications. Here is one to begin with.

Example ***(Arithmetic–geometric means inequality).*** Let $(x_i, 1 \leq i \leq n)$ be any collection of positive numbers and $(p_i; 1 \leq i \leq n)$ any collection of positive numbers such that $\sum_{i=1}^{n} p_i$. Show that

$$p_1x_1 + p_2x_2 + \cdots + p_nx_n \geq x_1^{p_1} x_2^{p_2} \cdots x_n^{p_n}. \tag{8.9}$$

Solution. Let X be the random variable with probability mass function $P(X = x_i) = p_i$; $1 \leq i \leq n$. Then, from the inequality $\log E(X) = \log(p_1x_1 + \cdots + p_nx_n) \geq E(\log X) = p_1 \log x_1 + \cdots + p_n \log x_n = \log(x_1^{p_1} x_2^{p_2} \cdots x_n^{p_n})$ the result (8.9) follows because $\log x$ is an increasing function.

In the special case when $p_i = 1/n$, $1 \leq i \leq n$, then (8.9) takes the form

$$\frac{1}{n}\sum_{i=1}^{n} x_i \geq \left(\prod_{i=1}^{n} x_i\right)^{\frac{1}{n}}.$$

8.3. The law of large numbers (The law of averages)

To begin with popular interpretation of the law of large numbers (LLN) it states that:

If the probability of a given outcome to an event is P and the event is repeated N times, then the larger N becomes, so the likelihood increases that the closer, in proportion, will be the occurrence of the given outcome to $N \times P$.

For example, if the probability of throwing a double 6 with two dice is 1/36, then the more times we throw the dice, the closer, in proportion, will be the number of double 6s thrown to of the total number of throws. This is what in everyday language is known as **the law of averages**. The overlooking of the vital words “in proportion” in the above definition leads to much misunderstanding.

The “**gambler's fallacy**” lies in the idea that “in the long run” chances will even out. Thus if a coin has been spun 100 times, and has landed 60 times head uppermost and 40 times tails, many gamblers will state that tails are now due for a run to get even. There are fancy names for this belief. The theory is called the “maturity” of chances, and the expected run of tails is known as a “corrective”, which will bring the total of tails eventually equal to the total of heads. The belief is that the law of averages really is a law which states that in the longest of long runs the totals of both heads and tails will eventually become equal.

In fact, the opposite is really the case. As the number of tosses gets larger, the probability is that the percentage of heads or tails thrown gets nearer to 50 %, but that the difference between the actual number of heads or tails thrown and the number representing 50 % gets larger.

Let us return to our example of 60 heads and 40 tails in 100 spins, and imagine that the next 100 spins result in 56 heads and 44 tails. The percentage of heads has now dropped from 60 % to 58 %. But there are now 32 more heads than tails, where there were only 20 before. The law of averages follower who backed tails is 12 more tosses to the bad. If the third hundred tosses result in 50 heads and 50 tails, there are now 166 heads in 300 tosses, down to approximately 55 %, but the tails backer is still 32 tosses behind.

Put another way, we would not be too surprised if after 100 tosses there were 60 % heads. We would be astonished if after a million tosses there were still 60 % heads, as we would expect the deviation from 50 % to be much smaller. Similarly, after 100 tosses, we are not too surprised that the difference between heads and tails is 20. After a million tosses we would be very surprised to find that the difference was not very much larger than 20.

A chance event is uninfluenced by the events which have gone before (we speak about independent events). If a true die has not shown 6 for 30 throws, the probability of a 6 is still 1/6 on the 31st throw.

It is interesting that despite significant statistical evidence and proof of all of the above people will go to extreme lengths to fulfill their belief in the fact that a “corrective” is due. The number 53 in an Italian lottery had failed to appear for some time and this lead to an obsession with the public to bet ever larger amounts on the number. People staked so much on this “corrective” that the failure of the number 53 to occur for two years was blamed for several deaths and bankruptcies.

In more precise and general mathematical setting we have following results.

Theorem 8.3 (Weak law of large numbers). *Let* $(X_n;\ n \geq 1)$ *be a sequence of independent random variables having the same finite mean and variance,* $\mu = E(X_1)$ *and* $\sigma^2 = \text{var}(X_1)$*. Then, as* $n \to \infty$,

$$(X_1 + \cdot \ldots \cdot + X_n)/n \xrightarrow{P} \mu.$$

It is customary to write $\overline{X}_n = n^{-1} \sum_{i=1}^{n} X_i$ (the **sample mean**).

Proof. Recall (8.4) inequality: for any random variable Y and $a > 0$, $P(|Y - E(Y)| \geq a) \leq \text{var}(Y)/a^2$. Hence, letting $S_n = \sum_{i=1}^{n} X_i$, $Y = (S_n - n\mu)/n$, we have $\forall\, \varepsilon > 0$ $P(|S_n - n\mu|/n > \varepsilon) \leq n^{-2} E(\sum_{i=1}^{n} (X_i - \mu))^2/\varepsilon^2 = n^{-2}\varepsilon^{-2} \sum_{i=1}^{n} \text{var}\,(X_i) = n^{-1}\varepsilon^{-2}\,\sigma^2 \to 0$ as $n \to \infty$, and the theorem is proved.

The law essentially states that for any nonzero margin specified, no matter how small, with a sufficiently large sample there will be a very high probability that the average of the observations will be close to the expected value, that is, within the margin.

An assumption of finite variance $\text{var}(X_1) = \text{var}(X_2) = \ldots = \sigma^2 < \infty$ is not necessary. Large or infinite variance will make the convergence slower, but the LLN holds anyway. This assumption is often used because it makes the proof easier and shorter.

The **strong law of large numbers** states that the sample average converges almost surely to the expected value as $n \to \infty$,

$$(X_1 + \cdot\ \cdot + X_n)/n \xrightarrow{\text{a.s.}} \mu.$$

That is, $P(\overline{X}_n \to \mu) = 1$.

This law justifies the intuitive interpretation of the expected value of a random variable as the "long-term average when sampling repeatedly".

Moreover, if the summands are independent but not identically distributed, then $\overline{X}_n - E(\overline{X}_n) \xrightarrow{a.s.} 0$ provided that each X_k has a finite second moment and $\sum_{k=1}^{\infty} k^{-2} \text{var}(X_k) < \infty$. This statement is known as **Kolmogorov's strong law**.

The weak law states that for a specified large n, the average $\overline{X}_n$ is likely to be near μ. Thus, it leaves open the possibility that the event $|\overline{X}_n - \mu| > \varepsilon$ happens an infinite number of times, although at infrequent intervals.

The strong law shows that this almost surely will not occur. In particular, it implies that with probability 1, we have that for any $\varepsilon > 0$ the inequality $|\overline{X}_n - \mu| < \varepsilon$ holds for all large enough n.

We can consider LLN as the linkage between the theoretical and experimental probabilities. Consider X_n in LLN as Bernoulli RVs and you receive that relative frequency $m/n \xrightarrow{\text{a.s.}/P} P(A)$ (the strong/weak LLN in Borel/Bernoulli forms respectively). Remember the statistical definition of probability ((1.2.1) formula).

Jakob Bernoulli (1659–1705) was the first to recognize the connection between long-run proportion and probability. In 1705, the year of his death, he provided a mathematical proof of the LLN in his book “Ars Conjectandi” (“The Art of Conjecturing”). This principle also plays a key role in the understanding of sampling distributions, enabling pollsters and researchers to make predictions based on statistics.

8.4. Sampling from a distribution. Central limit theorem

Performing the experiment will give us a single value of a random variable X. Repeating the experiment independently n times will give us the so called random **sample** of size n. The word “random” is usually omitted for the sake of brevity. The individual values X_1, X_2, ..., X_n are independent, identically distributed (iid) random variables (think about them as the would-be values, before the experiment is actually performed). A specific set of observed values (x_1, x_2, ..., x_n) is a set of sample values assumed by the sample.

The **sample mean** is (unlike the old “means” which were constant parameters) a random variable, defined by: $\overline{X}_n = \bar{X} = n^{-1}\sum_{i=1}^{n} X_i$. Its expected value $E(\bar{X}) = \mu$, where μ is the expected value of the distribution from which the sample is taken (sometimes also called “population” or “parent distribution”).

Similarly $\text{var}(\bar{X}) = \sigma^2/n$ where σ is the standard deviation of the original distribution. This implies that $\sigma(\bar{X}) = \sigma/\sqrt{n}$ (the standard deviation of $\bar{X}$ is $\sqrt{n}$ smaller than σ; sometimes it is also called the standard error of $\bar{X}$). Note that the standard error tends to zero as the sample size n increases.

So now we know how the mean and standard deviation of $\bar{X}$ relate to the mean and standard deviation of the population.

How about the shape of the $\bar{X}$ distribution, how does it relate to the shape of the original distribution? The surprising answer is: it doesn’t (for large n, in practice for n more than a handful, under some general conditions), instead, the distribution of $\bar{X}$ has always the same regular shape, common to many distributions, from which we may sample!

We already know that the mean and standard deviation of this distribution are μ and $\sigma/\sqrt{n}$ respectively, now we would like to establish its asymptotic (i.e. large n) shape. This is, in a sense, trivial: since $\sigma/\sqrt{n} \to 0$ as $n \to \infty$, we get in the $n \to \infty$ limit a degenerate (single-valued, with zero variance) distribution, with all probability concentrated at μ.

We can prevent this distribution from shrinking to a zero width by standardizing $\bar{X}$ first, i.e. defining a new random variable

$Z = (\bar{X} - \mu)/(\sigma/\sqrt{n})$ and investigating its asymptotic ($n \to \infty$) distribution instead (the new random variable has the mean of 0 and the standard deviation of 1, thus its shape cannot "disappear" on us).

We do this by constructing the MGF of Z and finding its $n \to \infty$ limit. Since $Z = \sum_{i=1}^{n}(X_i - \mu)/(\sigma/\sqrt{n})$ is the sum of independent, identically distributed random variables, its MGF is the MGF of $(X_1 - \mu)/(\sigma/\sqrt{n}) \equiv Y$, raised to the power of n.

We know that MGF $M_Y(t) = 1 + E(Y)t + E(Y^2)t^2/2 + E(Y^3)t^3/3! + ... = 1 + t^2/(2n) + st^3/(6n^{3/2}) + kt^4/(24n^2) + ...$ where $s, k,...$ is the skewness, kurtosis,… of the original distribution. Raising $M_Y(t)$ to the power of n and taking the $n \to \infty$ limit results in $e^{t^2/2}$ regardless of the values of $s, k, ...$ (assuming they exist), since they are divided by higher than one power of n.

Thus, we get a rather unexpected result: the distribution of Z has (for large n) the same symmetric shape (described by the above MGF limit), not in the least affected by the shape of the original distribution (from which the sample is taken).

At the end of the section we give a precise formulation of the central limit theorem in the simplest form.

The following theorem was proved in 1900–1901 by a Russian mathematician A.M. Lyapunov (1857–1918).

A.M. Lyapunov and A.A. Markov were contemporaries and close friends. A.M. Lyapunov considered himself as Markov's follower (although he was only a year younger). He made his name through Lyapunov's functions, a concept that proved to be very useful in analysis of convergence to equilibrium in various random and deterministic systems. A.M. Lyapunov died tragically, committing suicide after the death of his beloved wife, amid deprivation and terror during the civil war in Russia.

Theorem 8.4. *Suppose $X_1, X_2,..., X_n$, are independent, identically distributed random variables, with finite mean μ and variance σ^2.*

If $S_n = \sum_{i=1}^{n} X_i$, then $\forall\, y \in R^1$

$$\lim_{n\to\infty} P\left(\left(S_n - E(S_n)\right)/\sqrt{\mathrm{var}(S_n)} < y\right) =$$

$$= \lim_{n\to\infty} P\left(\left(\bar{X} - \mu\right)/(\sigma/\sqrt{n}) < y\right) = \Phi(y) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} \exp\left(-\frac{x^2}{2}\right) dx.$$

In fact, the convergence in previous equation is uniform in y.

Here, the CLT was stated for iid RVs, but modern methods can extend it to a much wider situation and provide an accurate bound on the speed of convergence.

Sometimes the CLT is called the **law of errors**.

The bean machine, also known as the quincunx or Galton box, is a device invented by Sir Francis Galton to demonstrate the law of errors and the normal distribution. See http://www.youtube.com/watch?v=AUSKTk9ENzg.

See also a good video about central limit theorem (empirical proof) http://www.youtube.com/watch?v=NUClFiP0Nhc&feature=related.

8.5. Exercises

1. Suppose that it is known that the number of items produced at a factory per week is a random variable X with mean 50.
(i) What can we say about the probability $X \geq 75$?
(ii) Suppose that the variance of X is 25. What can we say about $P(40 < X < 60)$?
2. Let X is binomial (4, 1/2). Use Chebyshov's inequality to estimate $P(|X - 2| \geq 2)$ and compare with the exact probability.
3. Let X_{10000} be the fraction of heads in 10,000 tosses. Use Chebyshov's inequality to bound $P(|X_n - 1/2| \geq 0.01)$ and the normal approximation to estimate this probability.
4. Let X have a Poisson distribution with mean 16. Estimate $P(X \geq 28)$ using (i) Chebyshov's inequality, (ii) the normal approximation.
5. Suppose that each of 300 patients has a probability of 1/3 of being helped by a treatment. Find approximately the probability that more than 120 patients are helped by the treatment.
6. A person bets you that in 100 tosses of a fair coin the number of heads will differ from 50 by 4 or more. What is the probability you will win this bet?
7. Suppose we toss a coin 100 times. Which is bigger, the probability of exactly 50 heads or at least 60 heads?
8. Suppose that 10 % of a certain brand of jelly beans is red. Use the normal approximation to estimate the probability that in a bag of 400 jelly beans there are at least 45 red ones.
9. To estimate the percent of voters who oppose a certain ballot measure, a survey organization takes a random sample of 200 voters. If 45 % of the voters oppose the measure, estimate the chance that (i) exactly 90 voters in the sample oppose the measure, (ii) more than half the voters in the sample oppose the measure.
10. A basketball player makes 80 % of his free throws on the average. Use the normal approximation to compute the probability that in 25 attempts he will make at least 23.
11. In a 162 game season find the approximate probability that a team with a 0.5 chance of winning will win at least 87 games.
12. Suppose we roll a die 600 times. What is the approximate probability that the number of 1's obtained lies between 90 and 110?
13. British Airways and United offer identical service on two flights from New York to London that leave at the same time. Suppose that they are competing for the same pool of 400 customers who choose an airline at random. What is the probability United will have more customers than its 230 seats?

14. An insurance company has 10,000 automobile policy-holders. The expected yearly claim per policyholder is $240 with a standard deviation of $800. Approximate the probability that the yearly claim exceeds $2.7 million.

15. On each bet a gambler loses $1 with probability 0.7, loses $2 with probability 0.2, and wins $10 with probability 0.1. Estimate the probability that the gambler will be losing after 100 bets.

16. Suppose we roll a die 10 times. What is the approximate probability that the sum of the numbers obtained lies between 30 and 40?

17. An airline knows that in the long run only 90 % of passengers who book a seat show up for their flight. On a particular flight with 300 seats there are 324 reservations. (i) Assuming passengers make independent decisions what is the chance that the flight will be over booked? (ii) Redo (i) assuming passengers travel in pairs and each pair flips a coin with probability 0.9 of heads to see if they will both show up or both stay home.

18. A student is taking a true/false test with 48 questions. (i) Suppose she has a probability $p = 3/4$ of getting each question right. What is the probability she will get at least 38 right? (ii) Answer the last question if she knows the answers to half the questions and flips a coin to answer the other half. Notice that in each case the expected number of questions she gets right is 36.

19. The number of students who enroll in a psychology class is Poisson with mean 100. If the enrollment is > 120 then the class will be split into two sections. Estimate the probability that this will occur.

20. A gymnast has a difficult trick with a 10 % chance of success. She tries the trick 25 times and wants to know the probability she will get exactly two successes. Compute the (i) exact answer, (ii) Poisson approximation, (iii) normal approximation.

21. Suppose that we roll two dice 180 times and we are interested in the probability that we get exactly 5 double sixes. Find (i) the normal approximation, (ii) the exact answer, (iii) the Poisson approximation.

22. A seed manufacturer sells seeds in packets of 50. Assume that each seed germinates with probability 0.99 independently of all the others. The manufacturer promises to replace, at no cost to the buyer, any packet with 3 or more seeds that do not germinate. (a) Use the Poisson to estimate the probability a packet must be replaced. (b) Use the normal to estimate the probability that the manufacturer has to replace more than 70 of the last 4000 packets sold.

23. A probability class has 30 students. As part of an assignment, each student tosses a coin 200 times and records the number of heads. What is the probability no student gets exactly 100 heads?

24. A die is rolled repeatedly until the sum of the numbers obtained is larger than 200. What is the probability that you can do this in 66 rolls or fewer?

25. Suppose that the checkout time at a grocery store has a mean of 5 minutes and a standard deviation of 2 minutes. Estimate the probability that a checker will serve at least 49 customers during her 4-hour shift.

26. A fair coin is tossed 2500 times. Find a number m so that the chance that the number of heads is between $1250 - m$ and $1250 + m$ is approximately 2/3.

27. Members of the Beta Upsilon Tau fraternity each drink a random number of beers with mean 6 and standard deviation 3. If there are 81 fraternity members, how much should they buy so that using the normal approximation they are 93.32 % sure they will not run out?

28. For a class project, you are supposed to take a poll to forecast the outcome of an election. How many people do you have to ask so that with probability 0.95 your estimate will not differ from the true outcome by more than 5 %?

29. Suppose we take a poll of 2,500 people. What percentage should the leader have for us to be 99 % confident that the leader will be the winner?

30. An electronics company produces devices that work properly 95 % of the time. The new devices are shipped in boxes of 400. The company wants to guarantee that k or more devices per box work. What is the largest k so that at least 95 % of the boxes meet the warranty?