# 1

# Probability Theoretic Preliminaries

The aim of this chapter is to present the definitions, formulae and results of probability theory we shall need in the main body of the book. Although we assume that the reader has had only a rather limited experience with probability theory and, if somewhat vaguely, we do define almost everything, this chapter is not intended to be a systematic introduction to probability theory. The main purpose is to identify the facts we shall rely on, so only the most important—and perhaps not too easily accessible—results will be proved. Since the book is primarily for mathematicians interested in graph theory, combinatorics and computing, some of the results will not be presented in full generality. It is inevitable that for the reader who is familiar with probability theory this introduction contains too many basic definitions and familiar facts, while the reader who has not studied probability before will find the chapter rather difficult.

There are many excellent introductions to probability theory: Feller (1966), Breiman (1968), K. L. Chung (1974) and H. Bauer (1981), to name only four. The interested reader is urged to consult one of these texts for a thorough introduction to the subject.

## 1.1 Notation and Basic Facts

A *probability space* is a triple $(\Omega, \Sigma, P)$, where $\Omega$ is a set, $\Sigma$ is a $\sigma$-field of subsets of $\Omega, P$ is a non-negative measure on $\Sigma$ and $P(\Omega) = 1$. In the simplest case $\Omega$ is a finite set and $\Sigma$ is $\mathscr{P}(\Omega)$, the set of all subsets of $\Omega$. Then $P$ is determined by the function $\Omega \to [0, 1], w \to P(\{w\})$, namely

$$P(A) = \sum_{w \in A} P(\{w\}), A \subset \Omega.$$

A real valued *random variable* (r.v.) $X$ is a measurable real-valued function on a probability space, $X : \Omega \to \mathbb{R}$.

Given a real valued r.v. $X$, its *distribution function* is $F(x) = P(X < x), -\infty < x < \infty$. Thus $F(x)$ is monotone increasing, continuous from the left, $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$. If there is a function $f(t)$ such that $F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$, then $f(t)$ is the *density function* of $X$. We say that a sequence of r.vs $(Y_n)$ *tends to* $X$ in *distribution* if $\lim_{n\to\infty} P(Y_n < x) = P(X < x) = F(x)$, whenever $x$ is a point of continuity of $F(x)$. The notation for convergence in distribution is $X_n \xrightarrow{d} X$. Of course, convergence in distribution depends only on the distributions of the r.vs in question.

If $h$ is any real-valued function on $\mathbb{R}$, the *expectation* of $h(X)$ is

$$E(h(X)) = \int_{\Omega} h(X)\mathrm{d}P = \int_{-\infty}^{\infty} h(x)\mathrm{d}F(x).$$

In particular, the *mean* of $X$, usually denoted by $\mu$, is $E(X)$ and the *nth moment* of $X$ is $E(X^n)$. Of course, these need not exist but, as they do exist for the r.vs we are going to consider, we shall assume that they exist. The *variance* of $X$ is $\sigma^2(X) = E\{(X-\mu)^2\} = E(X^2) - \mu^2$ and the standard deviation is the non-negative square root of this.

If $X$ is a non-negative r.v. with mean $\mu$ and $t > 0$, then

$$\mu \geq P(X \geq t\mu)t\mu.$$

Rewriting this slightly we get *Markov's inequality*:

$$P(X \geq t\mu) \leq 1/t. \tag{1.1}$$

Now let $X$ be a real-valued r.v. with mean $\mu$ and variance $\sigma^2$. If $d > 0$, then clearly

$$E\{(X-\mu)^2\} \geq P(|X-\mu| \geq d)d^2$$

so we have *Chebyshev's inequality*:

$$P(|X-\mu| \geq d) \leq \sigma^2/d^2. \tag{1.2}$$

As a special case of this inequality we see that if $\mu \neq 0$, then

$$P(X=0) \leq P(|X-\mu| \geq \mu) \leq \sigma^2/\mu^2. \tag{1.2$'$}$$

In fact, one can do a little better, for by the Cauchy inequality with $\Omega_0 = \{w : X(w) \neq 0\}$ we have

$$E(X)^2 = \left(\int_{\Omega_0} X\mathrm{d}P\right)^2 \leq \left(\int_{\Omega_0} X^2\mathrm{d}P\right)\left(\int_{\Omega_0} 1\mathrm{d}P\right) = E(X^2)\{1-P(X=0)\}.$$

Hence

$$P(X = 0) \le 1 - E(X)^2/E(X^2) = \sigma^2/(\mu^2 + \sigma^2). \tag{1.3}$$

Most of the r.vs we encounter are non-negative integer valued, so unless it is otherwise indicated (for example, by the existence of the density function), we assume that the r.v. takes only non-negative integer values. The distribution of such a r.v., $X$, is given by the sequence

$$p_k = P(X = k), k = 0, 1, \ldots.$$

Clearly $p_k \ge 0$ and $\Sigma_{k=0}^{\infty} p_k = 1$. Then the mean of $X$ is $\Sigma_{k=1}^{\infty} k p_k$ and the $n$th moment is $E(X^n) = \Sigma_{k=1}^{\infty} k^n p_k$. If $X, X_1, X_2, \ldots$ are non-negative integer valued r.vs then $X_n \xrightarrow{d} X$ if

$$\lim_{n\to\infty} P(X_n = k) = P(X = k) = p_k$$

for every $k$.

Write $\mathscr{L}(X)$ for the distribution (*law*) of a r.v. $X$. Given integer-valued r.vs $X$ and $Y$, the *total variation distance* of $\mathscr{L}(X)$ and $\mathscr{L}(Y)$ is

$$d(\mathscr{L}(X), \mathscr{L}(Y)) = \sup\{|P(X \in A) - P(Y \in A)| : A \subset \mathbb{Z}\}.$$

With a slight abuse of notation occasionally we write $d(X, Y)$ or $d(X, \mathscr{L}(Y))$ instead of $d(\mathscr{L}(X), \mathscr{L}(Y))$.

Clearly $X_n \xrightarrow{d} X$ iff $d(X_n, X) \to 0$. Of course, any information about the speed of convergence of $d(X_n, X)$ to 0 is more valuable than the fact that $X_n$ tends to $X$ in distribution.

Given a probability space $(\Omega, \Sigma, P)$ and a set $C \in \Sigma, P(C) > 0$, the *probability of a set* $A \in \Sigma$ *conditional on* $C$ is defined as

$$P(A|C) = P(A \cap C)/P(C).$$

Then $P_C = P(\cdot|C)$ is a probability measure on $(\Omega, \Sigma)$. A r.v. $X$ is said to be taken *condition on* $C$ if it is considered as a function on $(\Omega, \Sigma, P_C)$; the expectation of this new r.v., denoted by $E(X|C)$, is said to be the expectation of $X$ conditional on $C$.

Following Feller (1966) we use the notation $(x)_r = x(x-1)\ldots(x-r+1)$. Thus $(n)_n = (n)_{n-1} = n!$ and $\binom{n}{k} = (n)_k/(k)_k$. We define the *rth factorial moment* of a r.v. $X$ as $E_r(X) = E\{(X)_r\}$. Thus if

$$P(X = k) = p_k,$$

then

$$E_r(X) = \sum_{k=r}^{\infty} p_k (k)_r.$$

Note that if $X$ denotes the number of objects in a certain class then $E_r(X)$ is the expected number of *ordered r-tuples* of elements of that class.

The r.vs $X_1, X_2, \dots$ are said to be *independent* if for each $n$

$$P(X_i = k_i, i = 1, \dots, n) = \prod_{i=1}^{n} P(X_i = k_i)$$

for every choice of $k_1, k_2, \dots, k_n$.

Note that $E(X+Y) = E(X) + E(Y)$ always holds and if $X_1, X_2, \dots, X_n$ are independent, $E(X_i) = \mu_i$ and $E(X_i - \mu_i)^2 = \sigma_i^2$ then $E(\sum_i X_i) = \sum_i \mu_i$ and

$$\sigma^2\left(\sum_i X_i\right) = E\left[\left\{\sum_i (X_i - \mu_i)\right\}^2\right] = \sum_i \sigma_i^2.$$

In our calculations we shall often need the following rather sharp form of *Stirling's formula* proved by Robbins (1955):

$$n! = \left(\frac{n}{e}\right)^n \sqrt{2\pi n} e^{\alpha_n}, \tag{1.4}$$

where $1/(12n+1) < \alpha_n < 1/12n$.

Throughout the book we use Landau's notation $O\{f(n)\}$ for a term which, when divided by $f(n)$, remains bounded as $n \to \infty$. Similarly $h(n) = o\{g(n)\}$ means that $h(n)/g(n) \to 0$ as $n \to \infty$. Also, $h(n) \sim g(n)$ express the fact that $h(n)/g(n) \to 1$ as $n \to \infty$. Thus $h(n) \sim g(n)$ is equivalent to $h(n) - g(n) = o\{g(n)\}$. Note that a weak form of Stirling's formula (4) is $n! \sim (n/e)^n \sqrt{2\pi n}$. If the symbols $o$, $O$ or $\sim$ are used without a variable, then we mean that the relation holds as $n \to \infty$.

An immediate consequence of (1.4) is that if $1 \le m \le n/2$, then

$$\begin{aligned} e^{-1/(6m)} \frac{1}{\sqrt{2\pi}} \left(\frac{n}{m}\right)^m \left(\frac{n}{n-m}\right)^{n-m} \left(\frac{n}{m(n-m)}\right)^{1/2} &\le \binom{n}{m} \\ &\le \frac{1}{\sqrt{2\pi}} \left(\frac{n}{m}\right)^m \left(\frac{n}{n-m}\right)^{n-m} \left(\frac{n}{m(n-m)}\right)^{1/2}. \end{aligned} \tag{1.5}$$

On putting $p = m/n$ and $q = 1-p$ we find that if $m \to \infty$ and $n-m \to \infty$, then

$$\binom{n}{m} \sim (2\pi)^{-1/2} (p^p q^q)^{-n} (pqn)^{-1/2}.$$

By writing the binomial coefficients as quotients of products of factorials we obtain some other useful inequalities. If $b \le b + x < a$ and

$0 \le y < b \le a$, then

$$\left(\frac{a-b-x}{a-x}\right)^x \le \binom{a-x}{b}\binom{a}{b}^{-1} \le \left(\frac{a-b}{a}\right)^x \le e^{-(b/a)x} \tag{1.6}$$

and

$$\left(\frac{b-y}{a-y}\right)^y \le \binom{a-y}{b-y}\binom{a}{b}^{-1} \le \left(\frac{b}{a}\right)^y \le e^{-(1-b/a)y}. \tag{1.7}$$

For future reference we note some approximations of $\log(1+t)$. If $t > -1$, then

$$\log(1+t) \le \min\{t, t - \tfrac{1}{2}t^2 + \tfrac{1}{3}t^3\}, \tag{1.8}$$

if $t > 0$, then

$$\log(1+t) > t - \tfrac{1}{2}t^2, \tag{1.9}$$

if $0 < t < 0.45$, then

$$\log(1+t) > t - \tfrac{1}{2}t^2 + \tfrac{1}{4}t^3, \tag{1.9$'$}$$

if $0 < t < 0.69$, then

$$\log(1-t) > -t - t^2, \tag{1.10}$$

and if $0 < t < 0.431$, then

$$\log(1-t) > -t - \tfrac{1}{2}t^2 - \tfrac{1}{2}t^3. \tag{1.10$'$}$$

## 1.2 Some Basic Distributions

Throughout the book we shall keep to the convention that $0 < p < 1$ and $q = 1 - p$. We say that $X$ is a *Bernoulli r.v. with mean p* if $X$ takes only two values, 0 and 1, and $P(X = 1) = p$, $P(X = 0) = q$. Thus $X$ can be thought of as the outcome of tossing a biased coin, with probability $p$ of getting a head. Let $X^{(1)}, X^{(2)}, \dots$ be a sequence of independent Bernoulli r.vs with each $X^{(i)}$ having mean $p$. Then the r.v. $S_{n,p} = \sum_{i=1}^n X^{(i)}$ satisfies

$$P(S_{n,p} = k) = b(k; n, p) = \binom{n}{k} p^k q^{n-k},$$

and we say that $S_{n,p}$ has a *binomial distribution with parameters n and p*. By definition $b(k; n, p)$ is the probability that we get $k$ heads when tossing a coin $n$ times, provided the probability of getting a head is $p$. Since $E(X^{(i)}) = p$ and $\sigma^2(X^{(i)}) = E\{(X^{(i)} - p)^2\} = pq$, the binomial distribution with parameters $n$ and $p$ has mean $E(S_{n,p}) = pn$ and variance $\sigma^2(S_{n,p}) = pqn$.

It is easily seen that $b(k; n, p)$ is largest when $k$ approximates to $pn$, the expectation of $S_{n,p}$. Indeed, note that

$$\frac{b(k; n, p)}{b(k-1; n, p)} = 1 + \frac{(n+1)p - k}{kq}. \tag{1.11}$$

Hence, if $m$ is the unique integer satisfying $p(n+1)-1 < m \le p(n+1)$, then the terms $b(k; n, p)$ strictly increase up to $k = m-1$ and strictly decrease from $k = m$; furthermore, $b(m-1; n, p) < b(m; n, p)$ unless $m = p(n+1)$, when equality holds. Consequently from Stirling's formula (4) it follows that if $p$ is fixed then, as $n \to \infty$,

$$\max_k b(k; n, p) \sim \frac{1}{\sqrt{2\pi pqn}} = \frac{1}{\sigma\sqrt{2\pi}}.$$

Relations (1.11) and (1.5) imply the following simple but useful bound on the probability in the tail of the binomial distribution.

**Theorem 1.1** *Let $u > 1$ and $1 \le m = \lceil upn \rceil \le n-1$. Then*

$$\begin{aligned} P(S_{n,p} \ge upn) &= P(S_{n,p} \ge m) < \frac{u}{u-1} b(m; n, p) \\ &\le \frac{1}{\sqrt{2\pi}} \frac{u}{u-1} \left( \frac{n}{m(n-m)} \right)^{1/2} u^{-upn} \left( \frac{1-p}{1-up} \right)^{(1-up)n}. \end{aligned}$$

*Proof* If $m \le k-1$ then, by (1.11),

$$\frac{b(k; n, p)}{b(k-1; n, p)} \le \frac{(n-m)p}{(m+1)q} < \frac{1}{u},$$

so

$$\sum_{k=m}^{n} b(k; n, p) < \frac{u}{u-1} b(m; n, p).$$

To see the second inequality, for $1 < x < 1/p$ set

$$f(x) = x^{-px} \left( \frac{1-p}{1-px} \right)^{1-px}.$$

Then

$$\frac{\mathrm{d}}{\mathrm{d}x} \log f(x) = -p \log x - p(1-px) \log \frac{1-p}{1-px} < 0,$$

so putting $v = m/(pn)$, by (1.5) we find that

$$\begin{aligned} b(m;n,p) &\le \frac{1}{\sqrt{2\pi}}\left(\frac{n}{m(n-m)}\right)^{1/2}\left(\frac{n}{m}\right)^m\left(\frac{n}{n-m}\right)^{n-m} p^m(1-p)^{n-m} \\ &= \frac{1}{\sqrt{2\pi}}\left(\frac{n}{m(n-m)}\right)^{1/2} f(v)^n \\ &\le \frac{1}{\sqrt{2\pi}}\left(\frac{n}{m(n-m)}\right)^{1/2} f(u)^n. \qquad \square \end{aligned}$$

The binomial distribution describes the number of successes among $n$ trials, with the probability of a success being $p$. Now consider the number of failures encountered prior to the first success, and denote this by $Y$. Then clearly

$$P(Y = k) = q^k p, k = 0, 1, \ldots$$

The distribution defined above is said to be *geometric, with mean* $q/p$. It is easily checked that the mean is indeed $q/p$, the variance is $q/p^2$ and the $r$th factorial moment is $r!(q/p)^r$ (Ex. 3).

The number of failures prior to the $r$th success, say $Z_r$, is said to have a *negative binomial distribution*. The terminology is justified by the fact that

$$P(Z_r = k) = \binom{r+k-1}{k} p^r q^k, \quad k = 0, 1, \ldots$$

Since $Z_r$ is the sum of $r$ independent geometric r.vs, $E(Z_r) = rq/p$ and $\sigma^2(Z_r) = rq/p^2$.

A continuous version of the geometric distribution is the *exponential distribution* (or negative exponential distribution). A non-negative real valued r.v. $L$ is said to have an exponential distribution with parameter $\lambda > 0$ is

$$P(L < t) = 1 - e^{-\lambda t} \text{ for } t > 0.$$

The density function of this distribution is $\lambda\, e^{-\lambda t}$, and easy calculations show that $E(L) = 1/\lambda$ and $\sigma^2(L) = 1/\lambda^2$ (Ex. 4).

If $Y$ is a geometric r.v. with mean $(1-p)/p$, where $p > 0$ is small, then the distribution of $pY$ is close to the exponential distribution with mean 1 (Ex. 5).

The *hypergeometric distribution* with parameters $N$, $R$ and $n(0 < n < N, 0 < R < N)$ is defined by

$$\begin{aligned} q_k &= P(X = k) = \binom{R}{k}\binom{N-R}{n-k} \Big/ \binom{N}{n} \\ &= \binom{n}{k}\binom{N-n}{R-k} \Big/ \binom{N}{R}, \quad k = 0, \ldots, s, \end{aligned}$$

where $s = \min\{n, R\}$. Thus $q_k$ is the probability that if we select $n$ balls at random from a pool of $R$ red balls and $N - R$ blue balls then exactly $k$ of the balls will be red. If $n$ is fixed and, as $N \to \infty$, the ratio $R/N$ converges to a constant $p, 0 < p < 1$, then the hypergeometric distribution tends to the binomial distribution $b(k; n, p)$. This is an immediate consequence of the inequality

$$\binom{n}{k}\left(p - \frac{k}{N}\right)^k \left(q - \frac{n-k}{N}\right)^{n-k} < q_k < \binom{n}{k} p^k q^{n-k} \left(1 - \frac{k}{N}\right)^{-(n-k)}, \tag{1.12}$$

where $p = R/N$ and $q = 1 - p$. As we are interested mostly in the upper bound for $q_k$, we show the second inequality. Consider the second expression for $q_k$ and note that

$$\begin{aligned}\binom{N-n}{R-k} \Big/ \binom{N}{R} &= \frac{(R)_k (N-n)_{R-k}}{(N)_R} = \frac{(R)_k (N-R)_{n-k}}{(N)_n} \\ &\le \left(\frac{R}{N}\right)^k \left(\frac{N-R}{N-k}\right)^{n-k} = p^k q^{n-k} \left(1 - \frac{k}{N}\right)^{-n+k},\end{aligned}$$

since

$$\frac{N-R}{N-k} = \frac{N-R}{N} \frac{N}{N-k} = q\left(1 - \frac{k}{N}\right)^{-1}.$$

A r.v. $Y$ is said to have *Poisson distribution with mean* $\lambda > 0$ if

$$P(Y = k) = p(k; \lambda) = \mathrm{e}^{-\lambda} \lambda^k / k!, k = 0, 1, \ldots.$$

This distribution, which we shall denote by $P_\lambda$ or $P(\lambda)$, is also closely related to the binomial distribution. Indeed, if $\lambda = pn$ and $0 < p < 1$, then

$$b(k; n, p) = \frac{(n)_k p^k}{k!} \left(1 - \frac{\lambda}{n}\right)^{n-k} \le \frac{\lambda^k}{k!} \mathrm{e}^{\lambda(n-k)/n} = p(k; \lambda) \mathrm{e}^{pk}. \tag{1.13}$$

Also, if $\lambda = pn, 0 < p < \frac{1}{2}$ and $k \le n/2$, then, by (1.10),

$$\begin{aligned} b(k; n, p) &= \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^{n-k} \Pi_{i=1}^{k-1} \left(1 - \frac{i}{n}\right) \\ &\ge \frac{\lambda^k}{k!} \exp\left\{-\lambda - \lambda^2/n - \sum_{i=1}^{k-1} \left(\frac{i}{n} + \left(\frac{i}{n}\right)^2\right)\right\} \\ &\ge p(k; \lambda) \mathrm{e}^{-(\lambda^2 + k^2)/n}. \end{aligned} \tag{1.14}$$

As a trivial consequence of (1.13) and (1.14) we find that if $p$ depends on $n$ in such a way that $pn \to \lambda$ as $n \to \infty$, where $\lambda$ is a positive constant, then the binomial distribution with parameters $n$ and $p$ tends to the Poisson distribution with mean $\lambda$. In symbols: $S_{n,p} \xrightarrow{d} P_\lambda$.

The factorial moments of $S_{n,p}$ and $P_\lambda$ are calculated without the slightest difficulty:

$$E_r(S_{n,p}) = p^r(n)_r \text{ and } E_r(P_\lambda) = \lambda^r.$$

## 1.3 Normal Approximation

In probabilistic graph theory we often need good estimates for the probability in the tail of the binomial distribution. The best known example of such an estimate is the DeMoivre–Laplace limit theorem [see Feller (1966, pp. 174–195) or Rényi (1970a, pp. 204–210)]. We shall give it here together with some bounds valid for all values in a given range.

A random variable is said to be *normally distributed with mean* $\mu$ *and variance* $\sigma^2$ if it has density function

$$\frac{1}{\sigma\sqrt{2\pi}} e^{-(t-\mu)^2/2\sigma^2}.$$

This distribution is usually denoted by $N(\mu, \sigma)$. We shall reserve $\phi(t)$ and $\Phi(x)$ for the density function and distribution function of $N$ (0, 1), respectively:

$$\phi(t) = (t)\frac{1}{\sqrt{2\pi}} e^{-t^2/2} \text{ and } \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2} dt.$$

If $x > 0$ and $l \geq 0$, then (see Ex. 7)

$$1 + \sum_{m=1}^{2l+1} (-1)^m \frac{1 \cdot 3 \cdot \ldots \cdot (2m-1)}{x^{2m}} < \{1 - \Phi(x)\} \Big/ \left\{\frac{1}{x}\phi(x)\right\}$$
$$< 1 + \sum_{m=1}^{2l} (-1)^m \frac{1 \cdot 3 \cdot \ldots \cdot (2m-1)}{x^{2m}}. \quad (1.15)$$

In particular, as $x \to \infty$ we have

$$1 - \Phi(x) \sim \frac{1}{x}\phi(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{x} e^{-x^2/2}. \quad (1.15')$$

Our first aim is to give upper bounds for the terms $b(k; n, p)$ of the binomial distribution. The restrictions in the theorem below are for the sake of convenience.

**Theorem 1.2** *Suppose $pn \geq 1$ and $1 \leq hqn/3$. Then if $k \geq pn + h$, we have*

$$b(k; n, p) < \frac{1}{\sqrt{2\pi pqn}} \exp\left\{-\frac{h^2}{2pqn} + \frac{h}{qn} + \frac{h^3}{p^2n^2}\right\}.$$

*Proof*

We may and shall assume that $k = pn + h$. From inequality (1.5) we find

$$b(l; n, p) < \left(\frac{pn}{l}\right)^l \left(\frac{qn}{n-l}\right)^{n-l} [n/\{2\pi l(n-l)\}]^{1/2}$$

for every $l, 1 \leq 1 \leq n-1$. Hence

$$\begin{aligned}(2\pi pqn)^{1/2} b(k; n, p) &< \left(\frac{pn}{k}\right)^{k+1/2} \left(\frac{qn}{n-k}\right)^{n-k+1/2} \\ &= \left(\frac{pn}{pn+h}\right)^{pn+h+1/2} \times \left(\frac{qn}{qn-h}\right)^{qn-h+1/2} \\ &= \left(1 + \frac{h}{pn}\right)^{-pn-h-1/2} \left(1 - \frac{h}{qn}\right)^{-qn+h-1/2}.\end{aligned}$$

From inequalities (1.9) and (1.10′) we obtain that

$$\begin{aligned}(2\pi pqn)^{1/2} b(k; n, p) < \exp\Big\{&-(pn + h + \tfrac{1}{2})\left(\frac{h}{pn} - \frac{h^2}{2p^2n^2}\right) \\ &+ (qn - h + \tfrac{1}{2})\left(\frac{h}{qn} + \frac{h^2}{2q^2n^2} + \frac{h^3}{2q^3n^3}\right)\Big\}.\end{aligned}$$

The required inequality follows by expanding and simplifying the expression on the right and taking into account the conditions $pn \geq 1$ and $1 \leq h \leq qn/3$. □

Theorem 1.2 gives us the following upper bounds on the probability in the tail of the binomial distribution.

**Theorem 1.3** *With the assumptions in Theorem 1.2 we have*

$$P(S_{n,p} \geq pn + h) < \left(\frac{pqn}{2\pi}\right)^{1/2} \frac{1}{h} \exp\left\{-\frac{h^2}{2pqn} + \frac{h}{pqn} + \frac{h^3}{p^2n^2}\right\}.$$

*Proof* By relation (1.11), if $m \geq pn + h$ then

$$\frac{b(m+1; n, p)}{b(m; n, p)} \leq 1 - \frac{h+q}{q(pn+h+1)} = \lambda < 1.$$