MGF for Binomial (n,p) $\quad (1-p+pe^s)^n$

Example. Consider the sequence of random variable $X_n \sim \mathsf{Binomial(n, \lambda/n)}$. We can show that X_n converges in distribution to a $\mathsf{Poisson}(\lambda)$ random variable.

$$M_{X_n}(s) = \left(1-\frac{\lambda}{n}+\frac{\lambda}{n}e^s\right)^n = \left(1+\frac{\lambda}{n}(e^s-1)\right)^n$$

It is known that $\lim_{n\to\infty}(1+x/n)^n = e^x$. So

$$\lim_{n\to\infty} M_{X_n}(s) = \exp(\lambda(e^s-1))$$

which we can identify as the MGF of a Poisson distribution with mean λ.

Application: Run time of quicksort

Consider a list of n distinct numbers which we want to sort into increasing order. The quicksort algorithm begins by choosing an element from the list called the pivot. The pivot is compared with all other elements in the list and the list is then divided into two sublists:

- one comprising those elements less than the pivot,
- the other comprising those elements greater than the pivot.

This procedure is then repeated on the sublists until the entire list is sorted.

1st step require $(n-1)$ comparisons for list of length n.

(Initial list)	10	5	3	7	9	2	1
(With 7 as pivot)	5	3	2	1	**7**	10	9
(With 3 as pivot for left sub-list)	2	1	**3**	5	7	10	9
...							
(List sorted)	1	2	3	5	7	9	10

$(n-1)$ compar

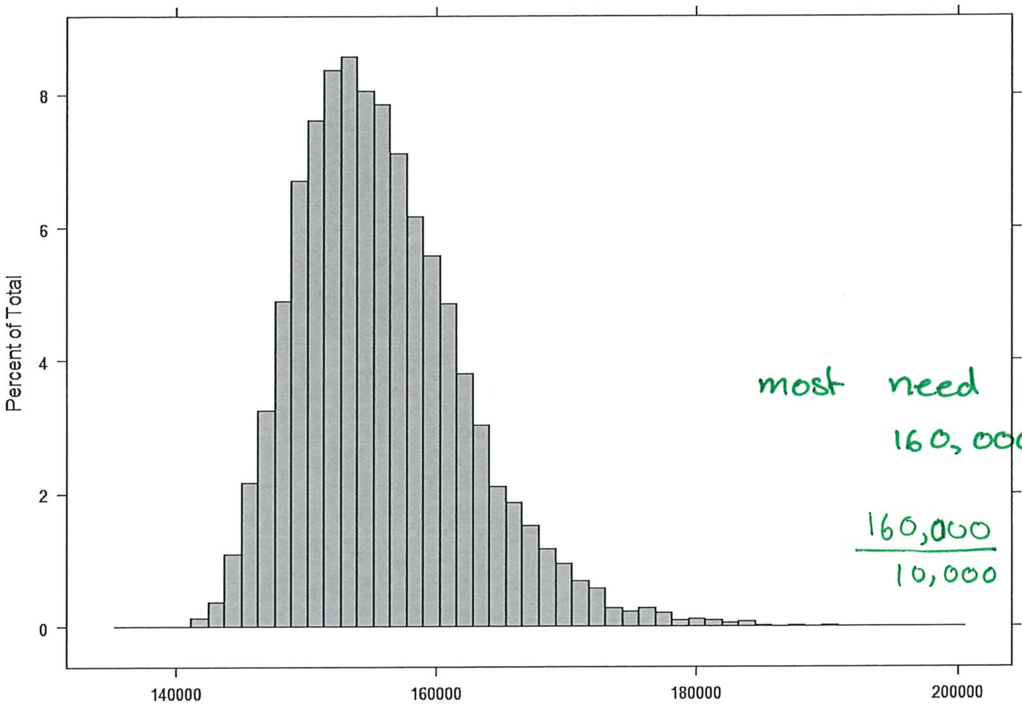
$(n-2)$ comp.

$(n-3)$ comp.

If the pivot is simply taken to be the first element of the list and the list is already sorted then all pairs of numbers will need to be compared. Therefore, the number of comparisons is $\boxed{n(n-1)/2}$. $\quad (n-1)+(n-2)+(n-3)+\cdots+2+1$ $\quad$ slow!

On the other hand, if the pivot is selected uniformly at random from the list, then the number of comparisons tends to be much smaller.

Let X_n be the random variable giving the number of comparison required to sort a list of n items. By convention $\mathbb{P}(X_0 = 0) = 1$. To compute $\mathbb{E}\,X_n$ we will condition of the selection of the first pivot



most need around 160,000 comparisons

$\frac{160,000}{10,000} \simeq 16$

Figure 4.7: Histogram of the number of comparisons made by quicksort when sorting random permutations of $\{1, 2, \ldots, 10000\}$.

$$\begin{aligned}
\mathbb{E}\, X_n &= \mathbb{E}\, [\mathbb{E}\, [X_n \,|\, \text{first pivot}]] \\
&= \sum_{k=1}^{n} \mathbb{P}(\text{1st pivot is } k^{th} \text{ largest})\, \mathbb{E}[X_n \,|\, \text{1st Pivot is } k^{th} \text{ largest}] \\
&= \sum_{k=1}^{n} \frac{1}{n} \times \left[(n-1) + \mathbb{E}[X_{k-1} + X_{n-k}]\right] * \\
&= (n-1) + \frac{1}{n} \sum_{k=1}^{n} \left(\mathbb{E}[X_{k-1}] + \mathbb{E}[X_{n-k}]\right)
\end{aligned}$$

Writing e_n for $\mathbb{E}\, X_n$ gives the recurrence equation

$$\begin{aligned}
e_n &= (n-1) + \frac{1}{n} \sum_{k=1}^{n} (e_{k-1} + e_{n-k}) \\
&= (n-1) + \frac{2}{n} \sum_{k=1}^{n} e_{k-1}
\end{aligned}$$

To solve this recurrence we need to do some re-arranging:

$$e_n = (n-1) + \frac{2}{n}\sum_{k=1}^{n} e_{k-1} \qquad \text{[multiply both sides by} n\text{]}$$

$$ne_n = n(n-1) + 2\sum_{k=1}^{n} e_{k-1} \qquad \text{[subtract } (n-1)e_{n-1} \text{ from both sides]}$$

$$ne_n - (n-1)e_{n-1} = n(n-1) + 2\sum_{k=1}^{n} e_{k-1} - \left\{(n-1)(n-2) + 2\sum_{k=1}^{n-1} e_{k-1}\right\}$$

$$= 2(n-1) + 2e_{n-1}$$

$$ne_n = (n+1)e_{n-1} + 2(n-1) \qquad \text{[divide by } n(n+1)\text{]}$$

$$\rightarrow \frac{e_n}{n+1} = \frac{e_{n-1}}{n} + \frac{2(n-1)}{n(n+1)} \qquad \text{[now iterating gives]}$$

$$\frac{e_n}{n+1} = 2\sum_{k=1}^{n} \frac{(k-1)}{k(k+1)} \qquad \text{[use partial fractions]}$$

$$= 2\sum_{k=1}^{n}\left\{\frac{2}{k+1} - \frac{1}{k}\right\}$$

$$= 2\sum_{k=1}^{n}\frac{1}{k} - \frac{4n}{n+1} \simeq 2\ln(n)$$

We note that $\ln n \leqslant \sum_{k=1}^{n} \frac{1}{k} \leqslant \ln n + 1$. To conclude, we see that

$$\lim_{n\to\infty} \frac{e_n}{2n\ln n} = 1.$$

$$\frac{e_n}{n+1} = \frac{e_{n-1}}{n} + \frac{2(n-1)}{n(n+1)}$$

$$= \frac{e_{n-2}}{n-1} + \frac{2(n-2)}{(n-1)n} + \frac{2(n-1)}{n(n+1)}$$

$$=$$

$$\frac{k-1}{k(k+1)} = \frac{A}{k} - \frac{B}{k+1} = \cancel{\frac{2}{k}} - \cancel{\frac{1}{k+1}} \quad \frac{2}{k+1} - \frac{1}{k}$$

$$= \frac{2k - 2(k+1)}{k(k+1)}$$

Nothing to see here.
Move along.

5

Continuous Random Variables

By the end of this chapter you should be able:

- To identify common continuous distributions.
- To compute the expectation and variance of a continuous random variable.
- To determine the probability density function of a transformed continuous random variable.
- To manipulate joint, marginal and conditional probability density functions.
- To compute the probability density function of a sum of two independent random variables using the convolution formula.
- To use moment generating functions to calculate moments and identify the distribution of a random variable.
- To understand the properties of the multivariate normal distribution.

So far we have dealt exclusively with discrete random variables. There are many situations where it may be more appropriate to use a continuous random variable. We will first recall the definitions of random variable and cumulative distribution function.

Definition. A function X assigning a real number to every outcome $\omega \in \Omega$ is called a **random variable**.

Definition. The **cumulative distribution function** (cdf) of X is the function $F : \mathbb{R} \to [0, 1]$ defined by

$$F(x) = \mathbb{P}(X \leqslant x) .$$

The following properties of the cumulative distribution function are basic consequences of the axioms of probability (Chapter 4):

- $0 \leqslant F(x) \leqslant 1$.
- $\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$.

- F is *non-decreasing*: If $x < y$, then $F(x) \leqslant F(y)$.
- F is *right-continuous*: If $x_n \downarrow x$, then $\lim_{n\to\infty} F(x_n) = F(x)$.

True for any random variable

We saw that for discrete random variables the cumulative distribution function is a step function. The cumulative distribution function of a continuous random variable has different behaviour.

Continuous Distributions

Definition. A random variable X is said to have a **continuous distribution** if its cumulative distribution function is continuous. A random variable with a continuous distribution is called a **continuous random variable**.

For a continuous random variable X and any $x \in \mathbb{R}$,

$$\begin{aligned}\mathbb{P}(X = x) &= \mathbb{P}(X \leq x) - \mathbb{P}(X < x) \\ &= F_X(x) - \lim_{y \uparrow x} F_X(y) \\ &= F_X(x) - F_X(x) = 0\end{aligned}$$

Therefore, if X is a continuous random variable, then $\boxed{\mathbb{P}(X \leqslant x) = \mathbb{P}(X < x)}$

Probability Density Function

For a continuous random variable X the probability $\mathbb{P}(X = x)$ is always 0. Hence, we cannot characterize the distribution of X via the probability mass function.

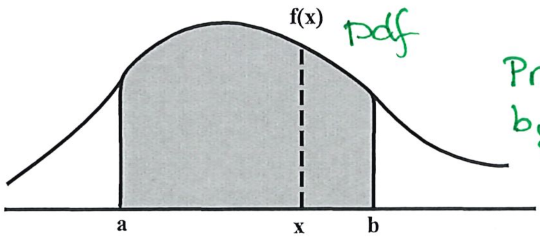
Instead, we have:

Definition. We say that a *continuous* random variable X has a **probability density function** (pdf) f_X if for all $x \in \mathbb{R}$

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\, du\ .$$

Hence, for all a, b

$$\mathbb{P}(a \leqslant X \leqslant b) = \int_a^b f_X(x)\, \mathrm{d}x. = \mathbb{P}(a < X < b)$$



Probability is determined by the area under the pdf.

By the Fundamental Theorem of Calculus if F_X is differentiable at x, then $\boxed{f_X(x) = F_X'(x)}$. $= \frac{d}{dx} F_X(x)$

We can interpret the probability density function $f_X(x)$ as the *"infinitesimal"* probability that $X = x$. More precisely, for small $h > 0$,

$$\mathbb{P}(x \leqslant X \leqslant x+h) = \int_x^{x+h} f_X(u)\, du \approx h\, f_X(x)\,.$$

But, note carefully, $f_X(x)$ is not a probability. In particular, it is not true that $f_X(x) = \mathbb{P}(X = x)$, for all x and we may have $f_X(x) > 1$ for some values of x.

Basic properties of f_X:

- $f_X(x) \geqslant 0$, for all x; — probabilities are non-negative
- $\int_{-\infty}^{\infty} f_X(x)\, dx = 1$ — $\mathbb{P}(\Omega) = 1$

Any function f satisfying these properties is the pdf of some distribution.

Important: Although not all continuous distributions have a probability density function, we will not consider those distributions in this course. Henceforth, when we discuss continuous distributions, we will assume the existence of a pdf.

Example: Let X be the random variable having pdf

$$f_X(x) = \begin{cases} c(1-x^2), & \text{if } x \in [-1,1] \\ 0, & \text{if } x \notin [-1,1]. \end{cases}$$

where c is a constant. What value must c take for f_X to be a valid pdf?

$$1 = \int_{-\infty}^{\infty} f_X(x)\, dx = \int_{-1}^{1} c(1-x^2)\, dx = c\left(x - \tfrac{1}{3}x^3\right)\Big|_{-1}^{1} = c\left(2 - \tfrac{2}{3}\right) = c \times \tfrac{4}{3}$$

Therefore, $c =$ $3/4$.

What is the appropriate notion of expectation for continuous random variables?

Definition. Let X be a continuous random variable with pdf f_X. Then, the expected value of X is defined as:

$$\mathbb{E}\,X = \int_{-\infty}^{\infty} u f_X(u)\, du.$$

In a way that is analogous to the definition of expectation for discrete random variables, $\mathbb{E}\,X$ is a weighted average of the values in the support of X, weighted, that is, according to the density f_X. We can also take expectations of functions of random variables.

Definition. Let X be a continuous random variable with pdf f_X and let g be any real-valued function. Then, the expected value of $g(X)$ is defined as:

$$\mathbb{E}\,g(X) = \int_{-\infty}^{\infty} g(u) f_X(u)\, du.$$

When we come to look at transformations, we will see that these two definitions are consistent. All of the properties of expectation mentioned in the previous chapter hold

(1) $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$.

$a, b \in \mathbb{R}$