

Appendix 3

Malcolm

Started 5 Nov 2024

Contents

A	Probability	2
A.1	Fundamental concepts	2
A.1.1	Permutations and Combinations, Binomial Coefficient . .	2
A.1.2	Expectation and Variance	4
A.1.3	Expected value of a function of a RV	5
A.1.4	Expectation and variance of linear functions	6
A.1.5	Variance in terms of Moments Expression	6
A.1.6	Expectation and Variance of Bernoulli	7
A.1.7	Expectation of Discrete Uniform	7
A.1.8	Variance of Discrete Uniform	9
A.2	Limit Theorems	10
A.2.1	Sample mean	10
A.2.2	Markov Inequality	11
A.2.3	Chebyshev Inequality	13
A.2.4	14
B	Supplementary Notes	15
B.1	The sum of the first n natural numbers is $n(n + 1)/2$	15
B.2	Telescoping series	16
B.3	Sum of series of products of consecutive integers	17
B.4	Sum of sequence of squares	18

Appendix A

Probability

A.1 Fundamental concepts

A.1.1 Permutations and Combinations, Binomial Coefficient

k-permutations

Starting with n distinct objects, and letting k be some positive integer where $k \leq n$, consider counting the number of different ways that we can pick k out of these n objects and arrange them into a sequence—the number of distinct k -object sequences.

We first have n choices for the first object. Having chosen the first, there are only $n - 1$ possible choices for the second, $n - 2$ for the third, and so on. This continues until we have chosen $k - 1$ objects, leaving us with $n - (k - 1)$ choices for the last one. The number of possible sequences, called *k-permutations*, can be written as

$$n(n - 1) \cdots (n - k + 1)$$

This can be rewritten, giving us

$$\begin{aligned} n(n - 1) \cdots (n - k + 1) &= \frac{n(n - 1) \cdots (n - k + 1)(n - k) \cdots 2 \cdot 1}{(n - k) \cdots 2 \cdot 1} \\ &= \frac{n!}{(n - k)!} \end{aligned}$$

See that in the special case where $k = n$ we have

$$n(n - 1)(n - 2) \cdots 2 \cdot 1 = n!$$

(This can also be seen from substituting $k = n$ into the formula and recalling the convention $0! = 1$.)

(next page)

Reordering a set

Starting with k objects, consider trying to find how many ways can we order them in a set of k elements. This follows a fairly similar principle to permutation; think of having k ‘slots’ to order k elements in: the first ‘slot’ has k possible inputs, the second $k - 1$ and so on. See that this just gives us $k!$.

Combinations

Combinations can be viewed as counting the number of k -element subsets of a given n -element set. Combinations are different from permutations in that *there is no ordering of selected elements*. For instance, where the 2-permutations of the letters A, B, C, and D are

AB, BA, AC, CA, AD, DA, BC, CB, BD, DB, CD, DC

the *combinations* of two out of these four letters are

AB, AC, AD, BC, BD, CD

See that the ‘duplicates’ are grouped together; for instance AB and BA are not viewed as distinct.

This reasoning can be generalised: each combination is associated with $k!$ ‘duplicate’ k -permutations—all ‘duplicate’ permutations of any given combination is just that permutation reordered for the maximum number of times:

(any single combination of length k) $\cdot k! =$ (permutations of that combination)

The number $n!/(n-k)!$ of k -permutations is equal to the number of combinations times $k!$. Hence the number of possible combinations is equal to

$$\frac{n!}{k!(n-k)!}$$

Binomial Coefficient

Consider a bernoulli process with probability p . We want the probability of k ‘successes’ in n trials. See that the probability of one *specific* sequence of n trials yielding k ‘successes’ would be

$$p^k(1-p)^{n-k}$$

We obtain the desired probability by multiplying this by the number of *combinations* of k ‘successes’ we can obtain in n trials:

$$\binom{n}{k} p^k (1-p)^{n-k}$$

(think tossing a coin three times and obtaining two heads—the heads might occur on the first and third tosses, or other *combinations* of trials).

A.1.2 Expectation and Variance

Expectation

We define the *expected value* of a random variable X with a PMF p_X by

$$\mathbb{E}[X] = \sum_x xp_X(x)$$

Variance and Standard Deviation

We define the *variance* associated with a random variable X as

$$\text{var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \sum_x (X - \mathbb{E}[X])^2 p_X(x)$$

(See that the because of the square the variance is always nonnegative). The variance provides a measure of dispersion of X around the mean. Another measure of dispersion is the *Standard deviation* of X , which is defined as the square root of the variance and is denoted by σ_X :

$$\sigma_X = \sqrt{\text{var}(X)}$$

The standard deviation is often easier to interpret because it has the same units as X .

A.1.3 Expected value of a function of a RV

Expectation of a function

Let X be a RV with PMF p_X , and let $g(X)$ be a function of X . Then the expected value of the random variable $g(X)$ is given by

$$\mathbb{E}[g(X)] = \sum_x g(x)p_X(x)$$

This can be shown, since

$$p_Y(y) = \sum_{\{x|g(x)=y\}} p_X(x)$$

we have

$$\begin{aligned}\mathbb{E}[g(X)] &= \mathbb{E}[Y] \\ &= \sum_y yp_Y(y) \\ &= \sum_y y \sum_{\{x|g(x)=y\}} p_X(x) \\ &= \sum_y \sum_{\{x|g(x)=y\}} yp_X(x) \\ &= \sum_y \sum_{\{x|g(x)=y\}} g(x)p_X(x) \\ &= \sum_x g(x)p_X(x)\end{aligned}$$

Variance

Using this we can write the variance of X as

$$\text{var}(X) = \mathbb{E}[(X - \mathbb{E}[X])^2] = \sum_x (X - \mathbb{E}[X])^2 p_X(x)$$

A.1.4 Expectation and variance of linear functions

We show for a random variable X , and letting $Y = aX + b$:

$$\boxed{\mathbb{E}[Y] = a\mathbb{E}[X] + b, \quad \text{var}(Y) = a^2\text{var}(X)}$$

Linearity of Expectations:

$$\mathbb{E}[Y] = \sum_x (ax + b)p_X(x) = a \underbrace{\sum_x xp_x(x)}_{=\mathbb{E}[X]} + b \underbrace{\sum_x p_x(x)}_{=1} = a\mathbb{E}[X] + b$$

Variance:

$$\begin{aligned} \text{var}(Y) &= \sum_x (ax + b - \mathbb{E}[aX + b])^2 p_X(x) \\ &= \sum_x (ax + b - a\mathbb{E}[X] + b)^2 p_X(x) \\ &= a^2 \sum_x (x - \mathbb{E}[X])^2 p_X(x) \\ &= a^2 \text{var}(X) \end{aligned}$$

Note that unless $g(X)$ is a linear function, it is not generally true that $\mathbb{E}[g(X)]$ is equal to $g(\mathbb{E}[X])$.

A.1.5 Variance in terms of Moments Expression

We show

$$\boxed{\text{var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2}$$

see that

$$\begin{aligned} \text{var}(X) &= \sum_x (x - \mathbb{E}[X])^2 p_X(x) \\ &= \sum_x (x^2 - 2x\mathbb{E}[X] + (\mathbb{E}[X])^2) p_X(x) \\ &= \sum_x x^2 p_X(x) - 2\mathbb{E}[X] \sum_x x p_X(x) + (\mathbb{E}[X])^2 \sum_x p_X(x) \\ &= \mathbb{E}[X^2] - 2(\mathbb{E}[X])^2 + (\mathbb{E}[X])^2 \\ &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \end{aligned}$$

A.1.6 Expectation and Variance of Bernoulli

Consider a Bernoulli RV X with PMF

$$p_X(k) = \begin{cases} p, & \text{if } k = 1. \\ 1 - p, & \text{if } k = 0. \end{cases}$$

The mean, second moment, and variance of X are as follows:

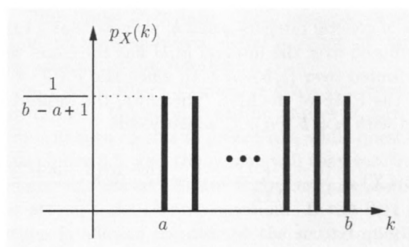
$$\begin{aligned} \mathbb{E}[X] &= 1 \cdot p + 0 \cdot (1 - p) = p \\ \mathbb{E}[X^2] &= 1^2 \cdot p + 0 \cdot (1 - p) = p \\ \text{var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = p - p^2 = p(1 - p) \end{aligned}$$

A.1.7 Expectation of Discrete Uniform

Consider a Discrete Uniform RV X with PMF, for $k \in [a, b]$:

$$p_X(k) = \begin{cases} \frac{1}{b-a+1}, & \text{if } k = a, a+1, \dots, b \\ 0, & \text{otherwise.} \end{cases}$$

An illustration is useful here:



Expectation

Upon inspection one might suppose that the expectation is

$$\mathbb{E}[X] = \frac{a+b}{2}$$

(next page)

Expectation (cont.)

The formula can be elucidated from the definition of the expectation. First see that a sequence $\sum_{k=a}^b k$ can be written as

$$\begin{aligned}\sum_{k=a}^b k &= \sum_{k=1}^b k - \sum_{k=1}^{a-1} k \\ &= \frac{(b)(b+1)}{2} - \frac{(a-1)(a)}{2} \quad (\text{see B.1}) \\ &= \frac{b^2 + b - a^2 + a}{2} = \frac{(b-a+1)(a+b)}{2}\end{aligned}$$

The last step isn't easy to factor, but working back from our 'hypothesis' for the expectation it coincides.

so now we have

$$\begin{aligned}\mathbb{E}[X] &= \sum_{k=a}^b k \left(\frac{1}{b-a+1} \right) \\ &= \frac{1}{b-a+1} \sum_{k=a}^b k \\ &= \frac{1}{b-a+1} \cdot \frac{(b-a+1)(a+b)}{2} \\ \mathbb{E}[X] &= \frac{(a+b)}{2}\end{aligned}$$

A.1.8 Variance of Discrete Uniform

Case for $k \in [1, n]$:

We can obtain the second moment for a discrete uniform distributed over $k \in [1, n]$ as

$$\begin{aligned}\mathbb{E}[X^2] &= \sum_{k=1}^n k^2 \left(\frac{1}{n}\right) \\ &= \frac{1}{n} \sum_{k=1}^n k^2 \\ &= \frac{1}{n} \cdot \frac{n(n+1)(2n+1)}{6} \quad (\text{see B.4}) \\ &= \frac{(n+1)(2n+1)}{6}\end{aligned}$$

We then use the formula for variance in terms of moments expression:

$$\begin{aligned}\text{var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \\ &= \frac{(n+1)(2n+1)}{6} - \left(\frac{n+1}{2}\right)^2 \\ &= \frac{1}{12}(n+1)(4n+2-3n-3) \\ &= \frac{n^2-1}{12}\end{aligned}$$

General case $k \in [a, b]$:

For the general case, note that a RV uniformly distributed over an interval $[a, b]$ has the *same variance* as one which is uniformly distributed over $[1, b-a+1]$ —the PMF of the second is just a shifted version of the PMF of the first.

Therefore, the desired variance is given by the first case, but instead with $n = b-a+1$, yielding

$$\boxed{\text{var}(X) = \frac{(b-a+1)^2-1}{12} = \frac{(b-a)(b-a+2)}{12}}$$

A.2 Limit Theorems

A.2.1 Sample mean

Definition

Here we discuss asymptomatic behaviour of sequences of random variables. The principal context involves a sequence X_1, X_2, \dots of independent identically distributed random variables with expectation μ and variance σ^2 . We denote

$$S_n = X_1 + \dots + X_n$$

to be the sum of the first n of them. Since they are independent we also have

$$\text{var}(S_n) = \text{var}(X_1) + \dots + \text{var}(X_n) = n\sigma^2$$

See that the distribution of S_n spreads out (it's variance increases) as n increases and doesn't have a meaningful limit. Consider instead the *sample mean*

$$M_n = \frac{X_1 + \dots + X_n}{n} = \frac{S_n}{n}$$

Expectation and Variance

We have the expectation as

$$\begin{aligned}\mathbb{E}[M_n] &= \frac{\mathbb{E}[X_1 + \dots + X_n]}{n} \\ &= \frac{\mathbb{E}[X_1] + \dots + \mathbb{E}[X_n]}{n} \\ &= \frac{n\mu}{n} = \mu\end{aligned}$$

and the variance as

$$\text{var}(M_n) = \frac{1}{n^2} \text{var}(S_n) = \frac{\sigma^2}{n}$$

See that the variance of M_n decreases to 0 as n increases.

With this consider a new random variable, that we modify based off M_n and S_n :

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

This has the properties

$$\mathbb{E}[Z_n] = 0, \quad \text{var}(Z_n) = \frac{\text{var}(S_n - n\mu)}{\sigma^2 n} = 1$$

A.2.2 Markov Inequality

Definition

Here we consider the *Markov inequality*. Loosely speaking it asserts that if a *nonnegative* random variable has a small mean, then the probability that it takes a large value must also be small:

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}, \quad \text{if } X \geq 0 \text{ and } a > 0.$$

(intuitively, as a increases, the probability that X is greater than it decreases)

Justification

Consider fixing a positive number a and considering the random variable Y_a defined by

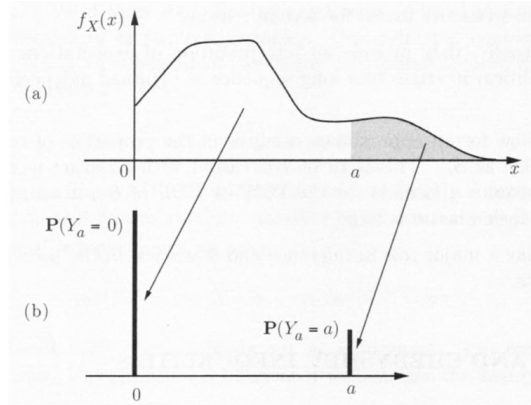
$$Y_a = \begin{cases} 0, & \text{if } X < a, \\ a, & \text{if } X \geq a. \end{cases}$$

See that the relation

$$Y_a \leq X$$

always holds and therefore

$$\mathbb{E}[Y_a] \leq \mathbb{E}[X]$$



See that all of the probability mass in the PDF of X between 0 and a is assigned to 0, and that above a assigned to a . Since mass is shifted to the left, the expectation can only decrease:

$$\mathbb{E}[X] \geq \mathbb{E}[Y_a] = a\mathbb{P}(Y_a = a) = a\mathbb{P}(X \geq a)$$

from which we obtain

$$a\mathbb{P}(X \geq a) \leq \mathbb{E}[X]$$

(next page)

Another justification

See that if $X \geq 0$ and $a > 0$:

$$\begin{aligned}\mathbb{E}[X] &= \int_0^\infty x f_X(x) \, dx \geq \int_a^\infty x f_X(x) \, dx \\ &\geq \int_a^\infty a f_X(x) \, dx \\ &= a \mathbb{P}(X \geq a)\end{aligned}$$

so

$$\mathbb{E}[X] \geq a \mathbb{P}(X \geq a)$$

and

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

A.2.3 Chebyshev Inequality

Definition

The *Chebyshev inequality*, loosely speaking, asserts that if a random variable has small variance, then the probability that it takes a value far from its mean is also small: Given a random variable X with mean μ and variance σ^2 ,

$$\mathbb{P}(|X - \mu| \geq c) \leq \frac{\sigma^2}{c^2}, \quad \text{for all } c > 0$$

Note that the Chebyshev inequality does not require the random variable to be negative.

Justification

Consider the nonnegative random variable $(X - \mu)^2$ and apply the Markov inequality with $a = c^2$ to obtain:

$$\mathbb{P}((X - \mu)^2 \geq c^2) \leq \frac{\mathbb{E}[(X - \mu)^2]}{c^2} = \frac{\sigma^2}{c^2}$$

Now observe that since the event $(X - \mu)^2 \geq c^2$ is identical to the event $|X - \mu| \geq c$, so that

$$\mathbb{P}(|X - \mu| \geq c) = \mathbb{P}((X - \mu)^2 \geq c^2) \leq \frac{\sigma^2}{c^2}$$

The Chebyshev inequality tends to be more powerful than the Markov inequality since it also uses information on the variance of X . An alternative form can also be obtained by letting $c = k\sigma$, $k > 0$, which yields

$$\mathbb{P}(|X - \mu| \geq k\sigma) \leq \frac{\sigma^2}{k^2\sigma^2} = \frac{1}{k^2}$$

(the probability that a random variable takes a value more than k standard deviations away from its mean is at most $1/k^2$)

Another justification

For a derivation that doesn't use the Markov inequality, introducing the function

$$g(x) = \begin{cases} 0, & \text{if } |x - \mu| < c, \\ c^2, & \text{if } |x - \mu| \geq c \end{cases}$$

since $(x - \mu)^2 \geq g(x)$ for all x we can write

$$\begin{aligned} \sigma^2 &= \int_{-\infty}^{\infty} (x - \mu)^2 f_X(x) dx \geq \int_{-\infty}^{\infty} g(x) f_X(x) dx \\ &= c^2 \left(\int_{-\infty}^{\mu-c} f_X(x) dx + \int_{\mu+c}^{\infty} f_X(x) dx \right) \\ &= c^2 \mathbb{P}(|X - \mu| \geq c) \end{aligned}$$

which can be arranged into the desired inequality.

A.2.4

Appendix B

Supplementary Notes

B.1 The sum of the first n natural numbers is $n(n+1)/2$

We have that

$$\sum_{i=1}^n i = 1 + 2 + \cdots + n$$

Now consider $2 \sum_{i=1}^n i$:

$$\begin{aligned} 2 \sum_{i=1}^n i &= 2(1 + 2 + \cdots + (n-1) + n) \\ &= (1 + 2 + \cdots + (n-1) + n) + (n + (n-1) + \cdots + 2 + 1) \\ &= (1 + n) + (2 + (n-1)) + \cdots + ((n-1) + 2) + (n + 1) \\ &= (n+1)_1 + (n+1)_2 + \cdots + (n+1)_n \\ &= n(n+1) \end{aligned}$$

so

$$\begin{aligned} 2 \sum_{i=1}^n i &= n(n+1) \\ \sum_{i=1}^n i &= \frac{n(n+1)}{2} \end{aligned}$$

B.2 Telescoping series

Let $\langle b_n \rangle$ be a sequence in \mathbb{R} . Let $\langle a_n \rangle$ be a sequence defined as

$$a_k = b_k - b_{k-1}$$

we show

$$\boxed{\sum_{k=m}^n a_k = b_n - b_{m-1}}$$

See that

$$\begin{aligned} \sum_{k=m}^n a_k &= \sum_{k=m}^n (b_k - b_{k-1}) \\ &= \sum_{k=m}^n b_k - \sum_{k=m}^n b_{k-1} \\ &= \sum_{k=m}^n b_k - \sum_{k=m-1}^{n-1} b_k \\ &= \left(\sum_{k=m}^{n-1} b_k + b_n \right) - \left(b_{m-1} + \sum_{k=m}^{n-1} b_k \right) \\ &= b_n - b_{m-1} \end{aligned}$$

B.3 Sum of series of products of consecutive integers

We show

$$\sum_{j=1}^n j(j+1) = 1 \cdot 2 + 2 \cdot 3 + \cdots + n(n+1) = \frac{n(n+1)(n+2)}{3}$$

See that

$$\begin{aligned} 3i(i+1) &= i(i+1)(i+2) - i(i+1)(i-1) \\ &= (i+1)((i+1)+1)((i+1)-1) - i(i+1)(i-1) \end{aligned}$$

Thus we have the basis of a telescoping series (see (B.2)):

$$3i(i+1) = b(i+1) - b(i)$$

where

$$b(i) = i(i+1)(i-1)$$

So we have

$$\begin{aligned} \sum_{j=1}^n 3j(j+1) &= \sum_{j=1}^n (j+1)((j+1)+1)((j+1)-1) - j(j+1)(j-1) \\ &= n(n+1)(n+2) - 0(0+1)(0-1) \\ &= n(n+1)(n+2) \end{aligned}$$

Thus

$$\sum_{j=1}^n j(j+1) = \frac{n(n+1)(n+2)}{3}$$

B.4 Sum of sequence of squares

We show

$$\forall n \in \mathbb{N} : \sum_{i=1}^n i^2 = \frac{n(n+1)(2n+1)}{6}$$

See that this follows from (B.3):

$$\begin{aligned} \sum_{i=1}^n 3i(i+1) &= n(n+1)(n+2) \\ \sum_{i=1}^n 3i^2 + \sum_{i=1}^n 3i &= n(n+1)(n+2) \\ \sum_{i=1}^n 3i^2 &= n(n+1)(n+2) - 3\frac{n(n+1)}{2} \quad \text{see (B.1)} \\ \sum_{i=1}^n i^2 &= \frac{n(n+1)(n+2)}{3} - \frac{n(n+1)}{2} \\ &= \frac{2n(n+1)(n+2) - 3n(n+1)}{6} \\ &= \frac{n(n+1)(2n+1)}{6} \end{aligned}$$