Chapter 5

Moments

5.1 Mean and variance

A random variable can be characterised in terms of the *shape* of the mass function. But usually we are interested in a small set of parameters that succinctly quantify the main features of the distribution.

Recall that, for a discrete random variable X, we defined the expectation of a function of a random variable X as:

$$\mathbb{E}\left(g(X)\right) \equiv \sum_i g(x_i) f_X(x_i) \qquad . \tag{5.1}$$

The 'kth moment' ($k = 0, 1, 2, \dots$) of X is defined as:

$$m_k \equiv \mathbb{E}\left(X^k\right) \qquad . \tag{5.2}$$

with the expectation (or mean) as simply the *first moment* ($k = 1$), very often denoted as μ.

$$m_1 = \mathbb{E}\left(X\right) = \mu \qquad . \tag{5.3}$$

The variance is defined in terms of the second moment as:

$$\mathrm{var}(X) \equiv \mathbb{E}\left(X^2\right) - \{\mathbb{E}\left(X\right)\}^2 = m_2 - m_1^2 \qquad . \tag{5.4}$$

and the positive square root of the variance is the *standard deviation*, denoted as σ:

$$\sigma \equiv \sqrt{\mathrm{var}\left(X\right)} \qquad . \tag{5.5}$$

and is a measure of the spread (dispersion) of the distribution.

Occasionally, but not in this particular course, higher-order moments are required, and it is useful to define the central moment of order k as:

$$\mu_k = \mathbb{E}\left((X - m_1)^k\right) \qquad . \tag{5.6}$$

The *skewness* (γ_1) and *kurtosis* (β_2) parameters are defined as follows:

$$\gamma_1 \equiv \frac{\mu_3}{\mu_2^{3/2}} \qquad , \qquad \beta_2 = \frac{\mu_4}{\mu_2} \qquad . \tag{5.7}$$

Skewness is the degree of asymmetry of a distribution, and is negative if values to the left (smaller X) of the mean have higher probabilities. Conversely, it has positive skewness when values to the right of the mean have higher probabilities.

Kurtosis is the degree of sharpness in the peak of a distribution, in particular with respect to the normal distribution for continuous variables.

Example

Show that $\mathbb{E}\left((X - m_1)^2\right) = \mathrm{var}\left(X\right)$.

Answer $g(X) = (X - m_1)^2$ thus:

$$\begin{aligned}
\mathbb{E}\left((X-m_1)^2\right) &= \sum_i (x_i - m_1)^2 f_X(x_i) = \sum_i (x_i^2 - 2m_1 x_i + m_1^2) f_X(x_i) && (5.8)\\
&= \left(\sum_i x_i^2 f_X(x_i)\right) - 2m_1 \left(\sum_i x_i f_X(x_i)\right) + m_1^2 \left(\sum_i f_X(x_i)\right) && (5.9)\\
&= \mathbb{E}\left(X^2\right) - 2m_1 m_1 + m_1^2 = m_2 - m_1^2 = \mathrm{var}(X). && (5.10)
\end{aligned}$$

Example

The discrete random variable $X = \{0, 1, 2, 3, \ldots, n\}$ has the Binomial distribution:

$$f_X(x) = \binom{n}{x} p^x q^{n-x} \quad , \quad x = 0, 1, \ldots, n \quad . \tag{5.11}$$

Calculate: $\mathbb{E}(X)$ and $\mathrm{var}(X)$.

According to the definition we have that:

$$\mathbb{E}(X) = \sum_{x=0}^{\infty} x \times \binom{n}{x} p^x q^{n-x} \quad . \tag{5.12}$$

In order to evaluate this finite series, we use a small 'trick' and the binomial theorem.

We note that, $xp^x = p\frac{\partial}{\partial p}(p^x)$, so we can write the expression as:

$$\mathbb{E}(X) = p\frac{\partial}{\partial p}\left(\sum_{n=0}^{n} \binom{n}{x} p^x q^{n-x}\right) \tag{5.13}$$

Now using the *binomial theorem*, namely:

$$(a+b)^n = \sum_{r=0}^{n} \binom{n}{r} a^r b^{n-r} \quad ,$$

the series can be summed immediately, to give:

$$\mathbb{E}(X) = p\frac{\partial}{\partial p}(p+q)^n = np(p+q)^{n-1} \quad . \tag{5.14}$$

Since, $p + q = 1$, this reduces to,

$$\mathbb{E}(X) = np \quad . \tag{5.15}$$

For the variance, we use a similar idea. Firstly, note the relation, $\mathbb{E}\left(X^2\right) = \mathbb{E}\left(X(X-1)\right) + \mathbb{E}(X)$. Next consider, $\mathbb{E}\left(X(X-1)\right)$, which can be written as:

$$\mathbb{E}\left(X(X-1)\right) = \sum_{x=0}^{n} x(x-1)\binom{n}{x} p^x q^{n-x} \quad . \tag{5.16}$$

Now, we can get rid of the awkward factor $x(x-1)$, with the finesse:

$$x(x-1)p^x = p^2\left(\frac{\partial^2}{\partial p^2}\right)p^x \quad .$$

Thus,

$$\mathbb{E}\left(X(X-1)\right) = p^2\left(\frac{\partial^2}{\partial p^2}\right)\sum_{x=0}^{n}\binom{n}{x} p^x q^{n-x} \quad . \tag{5.17}$$

Again the series is now in a familiar form and can be summed by the binomial theorem.

$$\mathbb{E}\left(X(X-1)\right) = p^2\left(\frac{\partial^2}{\partial p^2}\right)(p+q)^n \quad . \tag{5.18}$$

It is now time to evaluate the derivatives:

$$\mathbb{E}\left(X(X-1)\right) = p^2\left(\frac{\partial}{\partial p}\right)n(p+q)^{n-1} = p^2 n(n-1)(p+q)^{n-2} = n(n-1)p^2 \quad . \tag{5.19}$$

Finally, we gather terms together: $\mathbb{E}\left(X^2\right) = \mathbb{E}\left(X(X-1) + \mathbb{E}\left(X\right)\right.$, so that:

$$\text{var}(X) = \mathbb{E}\left(X^2\right) - \left(\mathbb{E}\left(X\right)\right)^2 = n(n-1)p^2 + np - (np)^2 = np(1-p) = npq. \tag{5.20}$$

A much simpler derivation of this result will be shown later.

5.2 Moment Generating Function

Moments can be calculated using the *moment generating function* defined as follows:

$$M_X(t) \equiv \mathbb{E}\left(e^{tX}\right) \quad . \tag{5.21}$$

Paradoxically, we rarely use the moment generating function to calculate moments - since there are usually easier ways to accomplish this task. However this function is extremely useful in other applications.

The *expectation* is nothing more than a series. If this series is absolutely convergent, and for most practical purposes this is the case, the sum can be differentiated as follows:

$$\frac{\partial}{\partial t}M_X(t) = \sum_i \frac{\partial}{\partial t}e^{tx_i}f_X(x_i) = \sum_i x_i e^{tx_i}f_X(x_i) \quad . \tag{5.22}$$

That is:

$$\frac{\partial}{\partial t}M_X(t) = M'_X(t) = \mathbb{E}\left(Xe^{tX}\right) \quad . \tag{5.23}$$

Taking this process to higher orders,

$$\left(\frac{\partial}{\partial t}\right)^n M_X(t) = M_X^{(n)}(t) = \mathbb{E}\left(X^n e^{tX}\right) \quad . \tag{5.24}$$

Now evaluating ths function at $t = 0$ leads to the important result:

$$\mathbb{E}\left(X^n\right) = M_X^{(n)}(0) \quad . \tag{5.25}$$

That is, we have an expression for the n moment in terms of the nth derivative of the generating function, and hence its name.

The *moment generating function* is very closely related to the Laplace transformation that arises in the theory of differential equations. Knowing how important the Laplace transform is, gives an appreciation as to the utility of the moment generating function.

Now critically, there is a *unique* (one-to-one) relationship between the mass function and moment generating function. This partly explains why it is useful, and for some purposes, it is more convenient to work with $M_X(t)$, rather than $f_X(x)$. Whatever we deduce to be true for the moment-generating function is true for the corresponding probability mass function. This will be apparent later in the discussion of the central-limit theorem and on of the applications in insurance and risk (ruin).

For the moment (sorry ... present), consider a simple illustration of how it can be used to calculate the mean and variance of the binomial distribution.

We have the mass function:

$$f_X(x) = \binom{n}{x}p^x q^{n-x} \qquad x = 0, 1, 2, \ldots, n \quad .$$

It follows that:

$$M_X(t) = \mathbb{E}\left(e^{tX}\right) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} e^{tx} \tag{5.26}$$

$$= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x q^{n-x} \tag{5.27}$$

Then, according to the binomial theorem, we obtain:

$$M_X(t) = (pe^t + q)^n \quad . \tag{5.28}$$

Now

$$M'_X(t) = npe^t(pe^t + q)^{n-1} \tag{5.29}$$

and so,

$$M'_X(0) = np \quad , \tag{5.30}$$

as before. But the second derivative is required for the variance, and this can be written, using the product rule as:

$$M''_X(t) = \left(\frac{\partial}{\partial t}\right)\left[npe^t(pe^t + q)^{n-1}\right] = npe^t(pe^t + q)^{n-1} + npe^t \, . \, (n-1)(pe^t + q)^{n-2}pe^t \tag{5.31}$$

that is:

$$M''_X(0) = p^2 n(n-1) + np \quad . \tag{5.32}$$

This is in agreement with the result calculated previously (equation 5.20). So that:

$$\text{var}\,(X) = p^2 n(n-1) + np - (np)^2 = np(1-p) = npq \quad . \tag{5.33}$$

Example

Calculate the mean, variance, and skewness of the Bernoulli distribution.

Solution

The probability mass function is:

$$f_X(x) = \begin{cases} q & x = 0 \\ p & x = 1 \end{cases} \tag{5.34}$$

with $q = 1 - p$, and $0 \leq p \leq 1$. The moment generating function is therefore:

$$M_X(t) = q + pe^t \tag{5.35}$$

Hence:

$$M'_X(t) = pe^t \quad \text{and} \quad M''_X(t) = pe^t \quad \text{and} \quad M'''_X(t) = pe^t \quad . \tag{5.36}$$

Thus:

$$m_1 = M'(0) = p \quad \text{and} \quad m_2 = p \quad \text{and} \quad m_3 = p \quad . \tag{5.37}$$

Therefore

$$m_1 = p \quad . \tag{5.38}$$

$$\text{var}\,(X) = \sigma^2 = m_2 - m_1^2 = p - p^2 = p(1-p) = pq \quad . \tag{5.39}$$

That is $\mu_2 = pq$ and:

$$\mu_3 = m_3 - 3m_2 m_1 + 3m_1 m_1^2 + m_1^3 = p - 3p^2 + 3p^3 - p^3 = p(q-p)q \tag{5.40}$$

$$\gamma_1 = \frac{p(q-p)q}{(pq)^{3/2}} = \frac{q-p}{\sqrt{qp}} \quad . \tag{5.41}$$

5.3 Probability Generating Function

It was stated that the *moment generating function* is analogous to the Laplace transform, beloved of physicists and engineers. Well another useful generating function, which corresponds to the Z–transform (almost as well-liked by electrical engineers) is defined as follows:

$$G_X(s) \equiv \mathbb{E}\left(s^X\right) = \sum_i f_X(x_i) s^{x_i} \qquad . \tag{5.42}$$

While the probability generating function is very useful in some applications, we don't encounter such applications in this course. However, it is worth mentioning just in case you encounter it in more advanced studies. Again, the skill in calculating this function is knowing how to sum series.

1. Binomial distribution

   According to the definition:

$$G_X(s) \equiv \mathbb{E}\left(s^X\right) = \sum_i s^{x_i} f_X(x_i) \qquad . \tag{5.43}$$

   so that, for $X = \{0, 1, \ldots, n\}$, and $q = 1 - p$:

$$G_X(s) = \sum_{x=0}^{n} \binom{n}{x} p^x q^{n-x} s^x = \sum_{x=0}^{n} \binom{n}{x} (ps)^x q^{n-x} \tag{5.44}$$

   Then using the binomial theorem, the sum can be evaluated:

$$G_X(s) = (ps + q)^n \qquad . \tag{5.45}$$

2. Poisson distribution

$$G_X(s) \equiv \mathbb{E}\left(s^X\right) = \sum_i f_X(x_i) s^{x_i} \qquad . \tag{5.46}$$

   so that, for $X = \{0, 1, 2, \ldots\}$:

$$G_X(s) = \sum_{x=0}^{\infty} e^{-\lambda} \frac{\lambda^x}{x!} s^x = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda s)^x}{x!} = e^{-\lambda} e^{\lambda s} = e^{\lambda(s-1)} \qquad . \tag{5.47}$$

5.4 Markov inequality

Consider a (positive) discrete random variable $X \geq 0$. Then, the following result, the *Markov inequality*, is true for any $a > 0$

$$\boxed{P(X \geq a) \leq \frac{\mu}{a} \qquad .} \tag{5.48}$$

To begin with, we note that,

$$P(X \geq a) \equiv \sum_{x_i \geq a} f_X(x_i) \qquad . \tag{5.49}$$

Then considering the mean (expected value):

$$\mu = \sum_{x_i} x_i f_X(x_i) = \sum_{x_i < a} x_i f_X(x_i) + \sum_{x_i \geq a} x_i f_X(x_i) \qquad . \tag{5.50}$$

This leads to the expression:

$$\mu - \sum_{x_i \geq a} x_i f_X(x_i) = \sum_{x_i < a} x_i f_X(x_i) \qquad . \tag{5.51}$$

Since the right-hand-side is clearly non-negative, one can write:

$$\mu - \sum_{x_i \geq a} x_i f_X(x_i) \geq 0 \qquad . \tag{5.52}$$

This can be written as, subtracting the same terms from either side:

$$\mu - \sum_{x_i \geq a} a f_X(x_i) \geq \sum_{x_i \geq a} (x_i - a) f_X(x_i) \qquad . \tag{5.53}$$

And again the right-hand side is non-negative and thus:

$$\mu \geq a \sum_{x_i \geq a} f_X(x_i) \qquad . \tag{5.54}$$

From this, and referring to (5.49), it immediately follows that:

$$P(X \geq a) \leq \frac{\mu}{a} \qquad , \tag{5.55}$$

which proves the result.

5.5 Expectations of functions

In general, $\mathbb{E}(f(X)) \neq f(\mathbb{E}(X))$. In fact the equality only arises for a linear function; $f(X) = aX + b$. However, there are some useful general inequalities that can be derived for particular classes of functions.

Suppose the function $g(X)$ has a special form - a convex or concave shape. Consider the function illustrated in the figure below (5.1). This is a convex (u-shaped) function in which the curvature (second derivative of the function) is always positive. That is, the function always exceeds the tangent. Below this in figure (5.2) we have an example of a concave function, for which the curvature is always negative.

More formally, a function $g(x)$ is *convex* on a closed interval $a \leq x \leq b$, if for any two points $x_{1,2} \in [a, b]$ and for any $0 < \lambda < 1$:

$$g(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda g(x_1) + (1 - \lambda) g(x_2) \tag{5.56}$$

That is any chord for the curve, a straight line joining two points $(x_1, g(x_1))$ and $(x_2, g(x_2))$, lies above the the curve points. So the point along the chord (line segment) with x-coordinate: $\lambda x_1 + (1 - \lambda)x_2$ $(0 < \lambda < 1)$ has a y-coordinate $\lambda g(x_1) + (1 - \lambda)g(x_2)$. This y-value is greater than the corresponding point on the curve with the same x-coordinate, hence the inequality!

The geometric interpretation of this inequality is easier to see. For a convex function, at the point $x = a$, the equation of the tangent is:

$$y(x) = g(a) + m(x - a) \tag{5.57}$$

with m the gradient of the line (which can be positive or negative). Thus we have:

$$g(x) \geq g(a) + m(x - a) \qquad . \tag{5.58}$$

For a concave function, $h(x)$, one in which the function always curves downwards, the opposite is true:

$$h(x) \leq h(a) + m(x - a) \qquad . \tag{5.59}$$

Again, a more formal definition of concave exists. A function $h(x)$ is *concave* on a closed interval $a \leq x \leq b$, if for any two points $x_{1,2} \in [a, b]$ and for any $0 < \lambda < 1$:

$$h(\lambda x_1 + (1 - \lambda)x_2) \geq \lambda h(x_1) + (1 - \lambda) h(x_2) \qquad . \tag{5.60}$$

Applying the inequality (5.58) to the random variable, $X = x$ (which can be either discrete or continuous), and letting $a = \mu = \mathbb{E}(X)$, then we have, for a convex function:

$$g(X) \geq g(\mathbb{E}(X)) + m(X - \mathbb{E}(X)) \quad . \tag{5.61}$$

Then taking expectations of both sides, and recalling that the probability mass/density is always positive and thus the direction of the inequality is not affected by such an operation, we arrive at *Jensen's inequality* for a *convex function*:

$$\boxed{\mathbb{E}(g(X)) \geq g(\mathbb{E}(X))} \quad . \tag{5.62}$$

Similarly, for a *concave function* we have:

$$\boxed{\mathbb{E}(h(X)) \leq h(\mathbb{E}(X))} \quad . \tag{5.63}$$

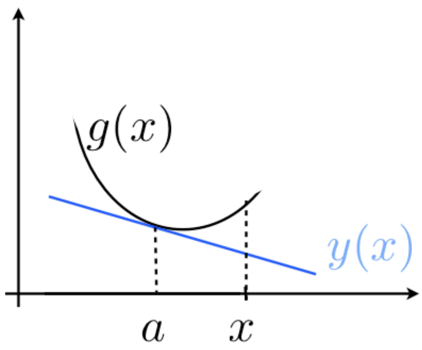

Figure 5.1: Illustration of a convex function, $g(x)$. Consider the tangent to the function $g(x)$ at the point $x = a$. Then it will have the equation: $y(x) = g(a) + m(x - a)$, where m is the gradient at $x = a$. Then the projection $y(x)$ will always lie below the curve: $y(x) \leq g(x)$, for any x.

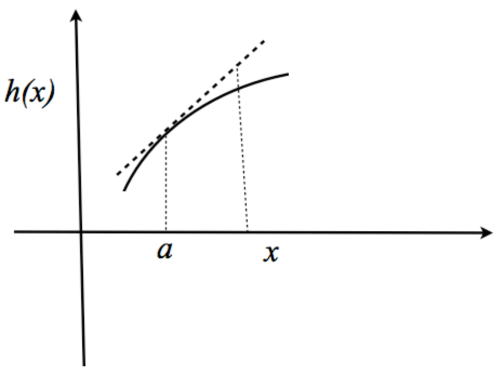

Figure 5.2: A simple concave function, $h(x)$. The tangent to $h(x)$ at the point $x = a$ has the equation: $y(x) = h(a) + m(x - a)$, where m is the gradient at $x = a$. Then the projection $y(x)$ will always lie below above the curve that is: $y(x) \geq h(x)$, for any x.

A particular case of a convex function would be the polynomial: $g(X) = X^n$, for $X \geq 0$ and $n > 1$, then we have:

$$\mathbb{E}(X^n) \geq (\mathbb{E}(X))^n \qquad n \geq 1 \quad . \tag{5.64}$$

which can also be written as:

$$\mathbb{E}(X) \leq (\mathbb{E}(X^n))^{1/n} \quad . \tag{5.65}$$

In particular:

$$\mathbb{E}\left(X^2\right) \geq \left(\mathbb{E}\left(X\right)\right)^2 \qquad , \tag{5.66}$$

which can also be shown from the Cauchy-Schwarz inequality. This gives the familiar result:

$$\mathbb{E}\left(X^2\right) - \left(\mathbb{E}\left(X\right)\right)^2 = \mathrm{var}(X) \geq 0 \qquad . \tag{5.67}$$

Similarly, if we take, $f(X) = |X|$, then the *triangle inequality*, familiar from linear algebra, appears in the guise:

$$\boxed{\mathbb{E}\left(|X|\right) \geq |\mathbb{E}\left(X\right)|} \quad . \tag{5.68}$$

Another useful result follows from the (convex) exponential function $g(x) = e^x$ in which case:

$$\mathbb{E}\left(e^X\right) \geq e^{\mathbb{E}(X)} \qquad , \tag{5.69}$$

with the equivalent form:

$$\mathbb{E}\left(X\right) \leq \ln \mathbb{E}\left(e^X\right) \qquad . \tag{5.70}$$

Another example of a concave function would be: $h(X) = X^\alpha$, where $X > 0$ and $0 < \alpha < 1$. Then:

$$\mathbb{E}\left(X^\alpha\right) \leq \left(\mathbb{E}\left(X\right)\right)^\alpha \qquad . \tag{5.71}$$

So if $\alpha = \frac{1}{2}$, then we have:

$$\mathbb{E}\left(X\right) \geq \left[\mathbb{E}\left(\sqrt{X}\right)\right]^2 \qquad . \tag{5.72}$$