Question. In the above example, what is the MGF $M_{\sum_{i=1}^n X_i}(t)$ when $\lambda_1 = \lambda_2 = \cdots = \lambda_n$?

$\left(\frac{\lambda}{\lambda-t}\right)^n, \quad t < \lambda.$

Exercise. Show that if $X_1 \sim \text{Normal}(\mu_1, \sigma_1^2)$ and $X_2 \sim \text{Normal}(\mu_2, \sigma_2^2)$ are independent,

$$aX_1 + bX_2 \sim \text{Normal}\big(a\mu_1 + b\mu_2,\ a^2\sigma_1^2 + b^2\sigma_2^2\big).$$

X_1 and X_2 have MGFs

$$M_{X_1}(t) = e^{\mu_1 t + \frac{\sigma_1^2 t^2}{2}} \text{ and } \quad M_{X_2}(t) = e^{\mu_2 t + \frac{1}{2}\sigma_2^2 t^2}$$

respectively. The MGF of $Y = aX_1 + bX_2$ is given by

$$\begin{aligned} M_Y(t) &= \mathbb{E}\left[e^{t(aX_1+bX_2)}\right] \\ &= \mathbb{E}\left[e^{taX_1}\right]\cdot\mathbb{E}\left[e^{tbX_2}\right] \quad \text{as } X_1 \text{ and } X_2 \text{ are independent.} \\ &= M_{X_1}(ta)\cdot M_{X_2}(tb) \\ &= \exp\left(\mu_1 ta + \tfrac{1}{2}\sigma_1^2 t^2 a^2\right)\cdot\exp\left(\mu_2 tb + \tfrac{1}{2}\sigma_2^2 t^2 b^2\right) \\ &= \exp\Big(\underbrace{(\mu_1 a + \mu_2 b)}_{\text{mean}} t + \tfrac{1}{2}\underbrace{(\sigma_1^2 a^2 + \sigma_2^2 b^2)}_{\text{variance}} t^2\Big) \end{aligned}$$

showing $Y \sim \text{Normal}\left(\mu_Y, \sigma_Y^2\right)$.

Whenever $M_X(t)$ is defined, it can be determined from the distribution of X. Is the converse true? Given a moment generating function $M_X(t)$, can the distribution of X be determined? Fortunately, the answer is YES (moment-generating functions have the **uniqueness property**) but proving it requires some complex analysis, so we won't do so in this class.

As a simple example, let's consider an easier task. Suppose that you know $M_X(t) = \mathbb{E}\,e^{tX}$ explicitly. Can you find the mean $\mathbb{E}\,X$ from this?

Assume we can interchange differentiation + integration.

$$\frac{d}{dt} M_X(t) = \mathbb{E}\left[\frac{d}{dt} e^{tX}\right] = \mathbb{E}\left[X e^{tX}\right] \Rightarrow M_X'(0) = \mathbb{E}[X]$$

More generally, given an MGF, we can find the nth moment $\mathbb{E}X^n$ by differentiating $M_X(t)$ n times and then setting $t = 0$. This is where the name *moment-generating function* comes from.

The series expansion $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!}$ (plus the linearity property of expectation) may be used to rewrite the MGF as:

$$M_X(t) = \mathbb{E}e^{tX} = \mathbb{E}\sum_{n=0}^{\infty} \frac{t^n X^n}{n!} = \sum_{n=0}^{\infty} \frac{t^n\,\mathbb{E}X^n}{n!}.$$

Note that a necessary (but it turns out, not sufficient) condition for the MGF to exist is that $\mathbb{E}\,X^n < \infty$ for all $n = 1, 2, \ldots$.

Exercise: Write $\mathbb{E}\,X$ and $\mathbb{E}\left[X^2\right]$ in terms of M_X.

$$M_X(t) =$$

$$\Rightarrow \frac{\mathrm{d}}{\mathrm{d}t} M_X(t) =$$

$$\Rightarrow M_X'(0) =$$

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2} M_X(t) =$$

$$\Rightarrow M_X''(0) =$$

$t < \lambda$, $\lambda > 0$

Exercise: Suppose X has moment generating function $M_X(t) = (\lambda/(\lambda - t))^n$, where n is a positive integer. Find the mean and variance of X.

$M(t) = (1 - t/\lambda)^{-n}$

$$M_X(t) = (1 - t/\lambda)^{-n}$$

$$M_X'(t) = -n(1 - t/\lambda)^{-(n+1)} \cdot (-1/\lambda) = \frac{n}{\lambda}(1 - t/\lambda)^{-(n+1)}$$

$$M_X''(t) = -\frac{n}{\lambda}(n+1)(1 - t/\lambda)^{-(n+2)} \cdot (-1/\lambda) = \cancel{\frac{n(n+1)}{\lambda^2}(1-t/\lambda)^{-(n+2)}}$$

$$= \frac{n(n+1)}{\lambda^2}(1 - t/\lambda)^{-(n+2)}$$

$$\mathbb{E}(X) = M_X'(0) = n/\lambda$$

$$\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}X)^2$$

$$= M_X''(0) - (M_X'(0))^2$$

$$= \frac{n(n+1)}{\lambda^2} - (n/\lambda)^2 = \frac{n}{\lambda^2}.$$

6

Estimation

<u>By the end of this chapter you should:</u>

- Know what a simple random sample is.
- Know the Law of Large Numbers and Central Limit Theorem.
- Be able to compute a confidence interval for a mean and for the difference of two means. parameter of normal distribution
- Be able to compute a confidence interval for a proportion and for the difference of two proportions. Parameter of Binomial distribution.

Collections of independent random variables play such an important role in statistics that we give them a special name.

Definition. The random variables $X_1, X_2, \ldots, X_n$ form **simple random sample** of size n if

(a) the X_i's are independent random variables, and

(b) every X_i has the same distribution.

We will often write $\mathbf{X}$ for the simple random sample $X_1, X_2, \ldots, X_n$.

A realisation of a simple random sample forms the *sample data.* We typically denote sample data as $\mathbf{x} = (x_1, x_2, \ldots, x_n)$.

Example: Let X_i be the face value of the ith roll of a fair, six-sided die. Assume that the rolls of the die are independent. The random vector (X_1, X_2, X_3) is a simple random sample of size 3 while an example of the sample data $\mathbf{x} = (x_1, x_2, x_3)$ might be $\mathbf{x} = (5, 1, 5)$.

We have seen a number of families of distributions. Each distribution in the family is determined by a set of parameters. For example, the normal distribution is parameterised by its $\boxed{\text{mean } (\mu)}$ and $\boxed{\text{variance } (\sigma^2)}$. We will denote the parameter of a generic distribution by θ. Typically, θ will be unknown and we will need to determine a suitable single number, based on sample data, that represents an appropriate value for θ.

Definition. A point estimate for θ is a single number $T(\mathbf{x})$ constructed from sample data $\mathbf{x}$ that can be thought of as a sensible value for θ. The random variable $T(\mathbf{X})$, where $\mathbf{X}$ is a random sample, is a point estimator of θ.

It is important to understand how well a given estimator performs. Some basic criteria for judging an estimator are *unbiasedness* and *consistency.*

Definition. We say $T(\mathbf{X})$ is an unbiased estimator of θ if $\mathbb{E}[T(\mathbf{X})] = \theta$. $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$

Example. The estimator $\bar{X}$ corresponding to the sample mean is an unbiased estimator of $\mu := \mathbb{E}[X_i]$ as

$$\boxed{\mathbb{E}[\bar{X}] = \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(X_i) = \frac{1}{n}\sum_{i=1}^{n} \mu = \mu.}$$

Example. The estimator S^2 corresponding to the sample variance

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

Try to show this.

is an unbiased estimator of $\sigma^2 := \mathrm{Var}(X_i)$. ~~We saw this in Quiz 7.~~

s.r.s

$X_i \sim Ber(p)$

$X = \sum_{i=1}^{n} X_i \sim Binom(n,p)$

Example. Let $X \sim \mathsf{Binomial}(n, p)$. The estimator $\widehat{P} := X/n$ is an unbiased estimator of p.

$$\boxed{\mathbb{E}\left[\widehat{P}\right] = \mathbb{E}[X/n] = \frac{1}{n}\mathbb{E}(X) = \frac{1}{n} \times np = p.}$$

Definition. The estimator $T(\mathbf{X})$ from a sample of size n is said to be consistent if, for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}(|T(\mathbf{X}) - \theta| > \varepsilon) = 0.$$

In order to establish consistency of our estimators, it will be useful to learn about the inequalities of Markov and Chebyshev (or is it Tchebycheff, or Tschebyschef, or ...)

Indicator function

$$\mathbb{I}(y \geqslant c) = \begin{cases} 1, & y \geqslant c \\ 0, & y < c \end{cases}$$

Inequalities of Markov and Chebyshev

Consider the function $g_c : [0, \infty) \to [0, \infty)$ defined by

$$g_c(y) := y - c\mathbb{I}(y \geqslant c). \geqslant 0$$

$$g_c(y) = \begin{cases} y - c, & y \geqslant c \\ y, & 0 \leqslant y < c \end{cases}$$

Let Y be any non-negative random variables. Then computing the expectation of $g_c(Y)$ shows

$$0 \leqslant \mathbb{E}[g_c(Y)] = \mathbb{E}[Y] - c\mathbb{E}[\mathbb{I}(Y \geqslant c)].$$

So

$$\boxed{\mathbb{E}[\mathbb{I}(Y \geqslant c)] = \mathbb{P}(Y \geqslant c) \leqslant \mathbb{E}[Y]/c}$$

Suppose $Z = \mathbb{I}(Y \geqslant c)$

$Z \sim$ Bernoulli

$$\mathbb{E}[Z] = \mathbb{E}(\mathbb{I}(Y \geqslant c)) = \int \mathbb{I}(y \geqslant c) f_Y(y)\,dy = \int_c^{\infty} f_Y(y)\,dy = \mathbb{P}(Y \geqslant c)$$

$Var(\bar{X}) = Var(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2} Var(\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n} Var(X_i) = \frac{n Var(X_1)}{n^2} = \frac{\sigma^2}{n}$

Recall $Var(X) = \mathbb{E}[(X-\mu)^2]$

This is called Markov's inequality. Now consider a random variable X with finite mean μ and variance σ^2. Setting $Y = (X - \mu)^2$ and $c = \varepsilon^2$ in Markov's inequality yields

$$\mathbb{P}(Y \geqslant c) \leqslant \mathbb{E}(Y)/c$$

$$\mathbb{P}(|X - \mu| \geqslant \varepsilon) = \mathbb{P}\left((X-\mu)^2 \geqslant \varepsilon^2\right) \leqslant \sigma^2/\varepsilon^2$$

This is called Chebyshev's inequality. We can use Chebyshev's inequality to show consistency of estimators. For example, let $\bar{X}$ be the estimator corresponding to the sample mean from a simple random sample of size n. Then

$\mathbb{E}[\bar{X}] = \mu$

$Var(\bar{X}) = \frac{\sigma^2}{n}$

$$\mathbb{P}(|\bar{X} - \mu| \geqslant \varepsilon)) \leqslant \frac{Var(\bar{X})}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2}$$

From this inequality we can see that $\bar{X}$ is a $\boxed{\text{Consistent}}$ estimator of μ. In other words, the average of a large number of independent and identically distributed random variables tends to the expected value as the sample size goes to infinity. This result is known as the **Law of Large Numbers**.

Central limit theorem

If $X_1, \ldots, X_n$ is a simple random sample from a $\mathsf{Normal}(\mu, \sigma^2)$ distribution, then

$$\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \sim$$

Suppose now we have a simple random sample $X_1, \ldots, X_n$ of size n, where $\mathbb{E}\,X_i = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$, but the distribution of the X_i is not necessarily normal. It is one of the remarkable results of probability and statistics that $\bar{X}$ has approximately a normal distribution with mean μ and variance σ^2/n. More precisely, for any $x \in \mathbb{R}$,

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{\bar{X} - \mu}{\sqrt{\sigma^2/n}} \leqslant x\right) = \Phi(x),$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. This result is called the **Central Limit Theorem**.

As a sketch of the ideas involved, suppose that $X_1, \ldots, X_n$ form a simple random sample with $\mathbb{E}\,X_i = 0$ and $\mathrm{Var}(X_i) = 1$. The general result follows by considering random variables $Y_i := \mu + \sigma X_i$. If we can show that the moment generating function $M_{Z_n}(t)$ of $Z_n := n^{-1/2}\sum_{i=1}^{n} X_i$ converges to $\exp(t^2/2)$ (the MGF of the standard normal distribution) for all t in some neighbourhood of 0, then it follows (*from Lévy's continuity theorem — a result we will not study*) that the distribution of Z_n converges to a standard normal distribution.

Let $M_X(t)$ denote the moment generating function of the X_i. The moment generating function of Z_n is