Probability and Distributions

1 Review

1.1 Conditional and unconditional expectations

The conditional expectation of X given Y is the expectation of X taken over the conditional probability distribution function (PDF):

$$\mathbb{E}[Y|X] = \begin{cases} \sum_y y f_{Y|X}(y|X), & \text{if } Y \text{ is discrete,} \\ \int_{-\infty}^{\infty} y f_{Y|X}(y|X) dy, & \text{if } Y \text{ is continuous.} \end{cases}$$

$f_{Y|X}(y|X)$ carries the **random variable** X as its argument, so the conditional expectation is also a random variable. However, we can also define the conditional expectation of Y given a particular value of X,

$$\mathbb{E}[Y|X=x] = \begin{cases} \sum_y y f_{Y|X}(y|x), & \text{if } Y \text{ is discrete,} \\ \int_{-\infty}^{\infty} y f_{Y|X}(y|x) dy, & \text{if } Y \text{ is continuous,} \end{cases}$$

which is just a number for any given value of x as long as the conditional density is defined.

Example (The Market for "Lemons"): The following is a simplified version of the famous paper by Akerlof (1970). Suppose that there are three types of X of used cars: excellent state (m for "melons"), average-quality (a), and bad condition (l for "lemons"). Each type of car is equally frequent, i.e.,

$$\Pr(X=m) = \Pr(X=a) = \Pr(X=l) = \frac{1}{3}.$$

The seller and buyer have the following (dollar) valuations Y_S and Y_B, respectively for each type of car:

Type	Seller	Buyer
Lemon	\$5,000	\$6,000
Average	\$6,000	\$10,000
Melon	\$10,000	\$11,000

Notice that for every type of car, the buyer's valuation is higher than the seller's, so for each type of car, trade should take place at a price between the buyer's and the seller's valuations. However, for used cars, quality is typically not evident at first sight, so if neither the seller nor the buyer know the type X of a car in question, their expected valuations are, by the law of iterated expectations (LIE),

$$\begin{aligned} \mathbb{E}[Y_S] &= \mathbb{E}[Y_S|X=l]\Pr(X=l) + \mathbb{E}[Y_S|X=a]\Pr(X=a) + \mathbb{E}[Y_S|X=m]\Pr(X=m) \\ &= \frac{1}{3}(5000+6000+10000) = 7000, \\ \mathbb{E}[Y_B] &= \mathbb{E}[Y_B|X=l]\Pr(X=l) + \mathbb{E}[Y_B|X=a]\Pr(X=a) + \mathbb{E}[Y_B|X=m]\Pr(X=m) \\ &= \frac{1}{3}(6000+10000+11000) = 9000, \end{aligned}$$

so trade should still take place.

But in a realistic setting, the seller of the used car knows more about its quality than the buyer (hidden information) and states a price at which he is willing to sell the car. If the seller can perfectly distinguish the three types of cars, whereas the buyer can't, the buyer should form expectations conditional on the seller willing to sell at the quoted price.

If the seller states a price less than \$6,000, the buyer knows for sure that the car is a "lemon" because otherwise the seller would demand at least \$6,000, i.e.,

$$\mathbb{E}[Y_B|Y_S < 6000] = \mathbb{E}[Y_B|X = l] = 6000,$$

and trade would take place. However, if the car was in fact a "melon", the seller would demand at least \$10,000, whereas the buyer would pay at most

$$\mathbb{E}[Y_B|Y_S \leq 10000] = \mathbb{E}[Y_B] = 9000 < 10000,$$

so that the seller won't be able to sell the high-quality car at a reasonable price.

The reason why the market for "melons" breaks down is that in this model, the seller can't credibly assure the buyer that the car in question is not of lower quality, so that the buyer factors the possibility of getting the bad deal into his calculation.

This leads us to our first proposition:

Proposition (Law of Iterated Expectations):

$$\mathbb{E}\left[\mathbb{E}[Y|X]\right] = \mathbb{E}[Y].$$

Proof: Let $g(x) = \mathbb{E}[Y|X = x]$, which is a function of x. We can now calculate the expectation

$$\begin{aligned}
\mathbb{E}[g(X)] &= \int_{-\infty}^{\infty} g(x) f_X(x) dx = \int_{-\infty}^{\infty} \mathbb{E}[Y|X = x] f_X(x) dx \\
&= \int_{-\infty}^{\infty} \left(\int_{\infty}^{\infty} y \frac{f_{XY}(x, y)}{f_X(x)} dy \right) f_X(x) dx \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{XY}(x, y) dy dx \\
&= \int_{-\infty}^{\infty} y \left(\int_{-\infty}^{\infty} f_{XY}(x, y) dx \right) dy \\
&= \int_{-\infty}^{\infty} y f_Y(y) dy \\
&= \mathbb{E}[Y]. \blacksquare
\end{aligned}$$

Proposition (Conditional Variance/Law of Total Variance):

$$\mathrm{Var}(Y) = \mathrm{Var}(\mathbb{E}[Y|X]) + \mathbb{E}[\mathrm{Var}(Y|X)],$$

which is known as the ANOVA identity. In particular, since $\mathrm{Var}(Y|X) \geq 0$, it follows that

$$\mathrm{Var}(Y) \geq \mathbb{E}[\mathrm{Var}(Y|X)],$$

which can, loosely speaking, be read as "knowing X decreases the variance of Y."

Proof: We can rewrite

$$\mathrm{Var}(\mathbb{E}[Y|X]) + \mathbb{E}[\mathrm{Var}(Y|X) = \left(\mathbb{E}\left[\mathbb{E}[Y|X]^2\right] - \left(\mathbb{E}\left[\mathbb{E}[Y|X]\right]\right)^2\right) + \left(\mathbb{E}\left[\mathbb{E}[Y^2|X]\right] - \mathbb{E}[\mathbb{E}[Y|X]^2\right),$$

where we use the property that $\mathrm{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$,

$$= \mathbb{E}[\mathbb{E}[Y|X]^2] - \mathbb{E}[Y]^2 + \mathbb{E}[Y^2] - \mathbb{E}[\mathbb{E}[Y|X]^2],$$

using the LIE

$$\begin{aligned} &= \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 \\ &= \mathrm{Var}(Y). \blacksquare \end{aligned}$$

1.2 Common distributions

Definition (Binomial Distribution): X is binomially distributed with parameters (n, p), $X \sim B(n, p)$ if its PDF is given by

$$f_X(x) = \begin{cases} \binom{n}{x} p^x (1-p)^{n-x}, & \text{if } x \in \{0, 1, ..., n\}, \\ 0, & \text{otherwise,} \end{cases}$$

with $\mathbb{E}[X] = np$ and $\mathrm{Var}(X) = np(1-p)$.

Definition (Uniform Distribution): X is uniformly distributed over the interval $[a, b]$, $X \sim U[a, b]$, it it has PDF

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & \text{if } a \le x \le b, \\ 0 & \text{otherwise.} \end{cases}$$

Definition (Exponential Distribution): X is exponentially distributed with parameter λ, $X \sim E(\lambda)$, if it has PDF

$$f_X(x) = \begin{cases} \lambda \exp(-\lambda x), & \text{if } x \ge 0, \\ 0 & \text{otherwise.} \end{cases}$$

Definition (Standard Normal Distribution): A random variable Z follows a standard normal distribution, $Z \sim N(0, 1)$, if its PDF is given by

$$f_Z(z) = \varphi(z) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{z^2}{2}\right),$$

for any $z \in \mathbb{R}$. Its CDF is denoted by

$$\Phi(z) = \Pr(Z \le z) = \int_{-\infty}^{z} \varphi(s) ds.$$

The CDF of a standard normal random variable doesn't have a closed-form expression (you have to look it up in tables or compute it).

1.2.1 Standardised random variable

Sometimes, it is useful to look at the following standardisation Z of a random variable X

$$Z = \frac{X - \mathbb{E}[X]}{\sqrt{\mathrm{Var}(X)}} = 0,$$

and

$$\mathrm{Var}(Z) = \frac{\mathrm{Var}(X - \mathbb{E}[X])}{\mathrm{Var}(X)} = \frac{\mathrm{Var}(X)}{\mathrm{Var}(X)} = 1.$$

For a standardised normal random variable, Z, $\mathbb{E}[Z] = 0$ and $\mathrm{Var}(Z) = 1$. It also turns out that binomial random variables are approximated by the normal for a large number n of trials:

Theorem (DeMoivre-Laplace Theorem): If $X \sim B(n,p)$ is a binomial random variable, then for any numbers $c \leq d$,

$$\begin{aligned}\lim_{n\to\infty} \Pr\left(c \leq \frac{X - np}{\sqrt{np(1-p)}} \leq d\right) &= \lim_{n\to\infty} \Pr\left(c < \frac{X - \mathbb{E}[X]}{\sqrt{\mathrm{Var}(X)}} \leq d\right) \\ &= \int_c^d \varphi_Z(z) dz,\end{aligned}$$

and notice that the transformation of the binomial variable to

$$Z = \frac{X - \mathbb{E}[X]}{\sqrt{\mathrm{Var}(X)}},$$

is in fact a standardisation . This result says that for large n, the probability that the standardised binomial random variable X falls inside the interval $(c, d]$ is approximately the same for a standard normal random variable.

1.2.2 Other nice properties of the normal distribution

For $Z \sim \mathcal{N}(0,1)$, we also call any random variable,

$$X = \mu + \sigma Z,$$

a normal random variable with mean μ and variance σ^2, i.e.,

$$X \sim \mathcal{N}(\mu, \sigma^2).$$

What is the PDF of X? By the change of variables formula, we have

$$\begin{aligned}f_X(x) &= \frac{1}{\sigma}\varphi\left(\frac{x-\mu}{\sigma^2}\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right).\end{aligned}$$

We can extend the same argument by nothing that the linear transformation of a normal random variable, X_1, is again normal.

Proposition: If $X_1 \sim \mathcal{N}(\mu, \sigma^2)$ and $X_2 = a + bX_1$, then

$$X_2 \sim \mathcal{N}(a + b\mu, b^2\sigma^2),$$

using the change of variables formula. We know that the expectation of the sum of n variables $X_1, ..., X_n$ is the sum of their expectations, and that the variance of n independent random variables is the sum of their variances. If the X_i's are also normal, then their sum is as well:

Proposition: If $X_1, ..., X_n$ are independent normal random variables with $X_i \sim \mathcal{N}(\mu_i, \sigma_i^2)$, then

$$Y = \sum_{i=1}^{n} X_i \sim \mathcal{N}\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right),$$

with PDF

$$f_Y(y) = \frac{1}{\sqrt{2\pi \sum_{i=1}^{n} \sigma_i^2}} \exp\left\{-\frac{\left(y - \sum_{i=1}^{n} \mu_i\right)^2}{2\sum_{i=1}^{n} \sigma_i^2}\right\}.$$

Proposition: If $Z_1, Z_2, ..., Z_k$ are independent with $Z_i \sim \mathcal{N}(0, 1)$, $Y = \sum_{i=1}^{k} Z_i^2$ is said to follow a χ^2 distribution with k degrees of freedom,

$$Y \sim \chi_k^2.$$

The expectation of a χ^2 distributed variable is given by the degrees of freedom:

$$\mathbb{E}[Y] = \sum_{i=1}^{k} \mathbb{E}[X_i^2] = k.$$

Proposition: If $X \sim \mathcal{N}(0, 1)$ and $Y \sim \chi_k^2$, then

$$T = \frac{Z}{\sqrt{Y}} = t_k,$$

i.e., T is follows the Student t-distribution with k degrees of freedom.

Proposition: If $Y_1 \sim \chi_{k_1}^2$ and $Y_2 \sim \chi_{k_2}^2$, then

$$F = \frac{\frac{Y_1}{k_1}}{\frac{Y_2}{k_2}} \sim F(k_1, k_2),$$

is said to follow the Fisher-Snedecor F-distribution with (k_1, k_2) degrees of freedom.

2 Probability densities

Find a table of tail probabilities for the normal distribution. Let $X \overset{d}{\sim} N(0, 1)$.

2.1

What is $P(X > 1.96)$*?*

$$
\begin{aligned}
P(X > 1.96) &= 1 - P(X \leq 1.96) \\
&= 1 - 0.975 \\
&= 0.025.
\end{aligned}
$$

What is $P(X > 2)$*?*

$$
\begin{aligned}
P(X > 2) &= 1 - P(X \leq 2) \\
&= 1 - 0.9772 \\
&= 0.0228.
\end{aligned}
$$

What is $P(|X| < 2)$*?*

$$
\begin{aligned}
P(|X| < 2) &= P(0 < X < 2) + P(-2 < X < 0) \\
&= (0.9772 - 0.5) + (0.5 - 0.02275) \\
&= 0.9544.
\end{aligned}
$$

2.2

Find q *so that* $P(|X| > q) = p$*. For* $p = 0.1$:

Since the normal distribution is symmetric, we need to find a one-sided density of $\frac{p}{2} = 0.05$, and its corresponding random variable. For the standard normal distribution this is approximately -1.64:

$$
\begin{aligned}
P(X < -1.64) &= 0.505, \\
\therefore P(|X| > 1.64) &\approx 0.1.
\end{aligned}
$$

For $p = 0.05$:

$$
\begin{aligned}
P(X < -1.96) &= 0.025, \\
\therefore P(|X| > 1.96) &= 0.05.
\end{aligned}
$$

For $p = 0.01$:

$$
\begin{aligned}
P(X < -2.58) &= 0.00494, \\
\therefore P(|X| > 2.58) &\approx 0.01.
\end{aligned}
$$

2.3

What is $P(X > 1.96)$ *when* $X \overset{d}{\sim} N(2, 1)$?

The probability density for $P(X > 1.96)$ when $X \overset{d}{\sim} N(2, 1)$ is analogous to the case of $P(X > -0.04)$ when $X \overset{d}{\sim} N(0, 1)$, since we simply have a shift in mean but retain unit variance. This density is $1 - 0.48405 = 0.51595$.

What is $P(X > 1.96)$ when $X \overset{d}{\sim} N(2, 4)$?

We need to make an adjustment due to the scale parameters on both the mean and variance. We know that a standard normal distribution with scale parameters μ and σ has density

$$\Phi\left(\frac{X - \mu}{\sigma}\right).$$

In this instance, for $\mu = 2$ and $\sigma = 2$, we have

$$\Phi\left(\frac{1.96 - 2}{2}\right) = \Phi(-0.02),$$
$$\therefore P(X > -0.02) = 1 - 0.49202 = 0.500798.$$

3 The Bernoulli distribution

Let $Y_1, ..., Y_n$ be independent Bernoulli(θ). Consider estimators $\hat{\theta} = \bar{Y}$ and $\tilde{\theta} = Y_n$.

3.1

Show that $\mathbb{E}(\hat{\theta}) = \mathbb{E}(\tilde{\theta}) = \theta$. We say that $\hat{\theta}$ and $\tilde{\theta}$ are unbiased estimators.

Given $Y \overset{d}{\sim} \text{Bern}(\theta)$, $\hat{\theta} = \bar{Y}$, $\tilde{\theta} = Y_n$. The probability mass function (PMF) is given as $\theta^y(1-\theta)^{1-y}$, $y = 0, 1$. This implies that for the n-th draw of Y, the expected value is given by:

$$\begin{aligned}
&\mathbb{E}(Y) = 0 \times \theta^0(1-\theta)^{1-0} + 1 \times \theta^1(1-\theta)^{1-1} = \theta, \\
\implies &\mathbb{E}(\tilde{\theta}) = \mathbb{E}(Y_n) = \theta.
\end{aligned}$$

The mean of Y, $\bar{Y}$, i.e. the maximum likelihood (ML) estimator for θ is given as:

$$\hat{\theta} = \bar{Y} = \frac{1}{n}\sum Y_i,$$

since

$$\begin{aligned}
\mathbb{E}(\hat{\theta}) &= \mathbb{E}\left(\frac{1}{n}\sum Y_i\right) \\
&= \frac{1}{n}\sum E(Y_i) = \frac{1}{n}\sum \theta \\
&= \frac{1}{n} n\theta = \theta,
\end{aligned}$$

where we can go from the first to the second line due to the linearity of expectations.

3.2

Compare the variances of $\hat{\theta}$ and $\tilde{\theta}$. Which has smaller variance? When comparing two unbiased estimators, the estimator with smaller variance is said to be more efficient.

The variance for $\hat{\theta}$ is given by:

$$\text{Var}(\hat{\theta}) = \frac{\theta(1-\theta)}{n},$$

since

$$\text{Var}(\hat{\theta}) = \bar{Y} = \frac{1}{n^2}\sum \text{Var}(Y_i) = \frac{\text{Var}(Y_i)}{n} = \frac{\theta(1-\theta)}{n},$$

because we assumed independence (otherwise we would also have covariances). The variance for $\tilde{\theta}$ is:

$$\text{Var}(\tilde{\theta}) = \text{Var}(Y_n) = \mathbb{E}(Y_n^2) - \mathbb{E}(Y_n)^2 = \theta(1-\theta).$$

We can clearly see that the ML estimator is more efficient as its variance tends to zero asymptotically.

4 Continuous distributions

In this exercise we will illustrate the concepts for continuous distributions introduced so far. Calculations with the normal distribution are difficult. Instead, we will look at the exponential distribution. It has distribution function

$$F_X(x) = 1 - \exp(-x),\ x > 0.$$

4.1

Argue that F_X is a distribution function, that is $F_X(0) = 0$, $\lim_{x\to\infty} F_X(x) = 1$, and F_X is non-decreasing. Is F_X continuous?

First, we know that $F(0) = 0$, and that

$$\lim_{x\to\infty} F(X) = \lim_{x\to\infty} [1 - \exp(-x)] = 1.$$

$F(X)$ is non-decreasing as $\frac{dF(X)}{dx} = \exp(-x) > 0$, and it is right continuous as the function is defined over $\mathbb{R}^+$. In other words, $\lim f(x_n) = f(\lim x_n)$. More formally,

$$\begin{aligned}
\lim_{h\to 0+} F_X(x+h) &= \lim_{h\to 0+} [1 - \exp(-(x+h))] \\
&= \lim_{h\to 0+} [1 - \exp(-x)\exp(-h)] \\
&= 1 - \exp(-x) \underbrace{\lim_{h\to 0+} \exp(-h)}_{=1} \\
&= 1 - \exp(-x) \\
&= F_X(x).
\end{aligned}$$

Therefore, F_X is a distribution function.

4.2

Find density of X.

If

$$F(X) = \int^x f(x)dx,$$

then

$$\frac{dF(X)}{dx} = f(x),$$

so

$$\frac{d}{dx}(1 - \exp(-x)) = \exp(-x) = f(x).$$

4.3

Find density of $Y = \lambda X$ using the change of variable technique.

First, we find the inverse function:

$$x = v(y) = \frac{y}{\lambda},$$

then take the derivative with respect to y to get $v'(y) = \frac{1}{\lambda}$. Then, apply the change of variable formula:

$$f_Y(y) = f_X(v(y)).v'(y) = \exp\left(-\frac{y}{\lambda}\right).\frac{1}{\lambda} = \frac{\exp(\frac{-y}{\lambda})}{\lambda}.$$

4.4

Show that $\mathbb{E}[X] = Var(X) = 1$.

To show that the expected value and variance are equal to unity:

$$\begin{aligned}
\mathbb{E}(X) &= \int_0^\infty x\exp(-x)dx, \\
&= [-x\exp(-x)]_0^\infty + \int_0^\infty \exp(-x)dx, \\
&= \int_0^\infty \exp(-x)dx = 1.
\end{aligned}$$

Recall that integration by parts is

$$\int_a^b u\ dv = [uv]_a^b - \int_a^b v\ du,$$

where we choose u by the following order: LIATE: Logs, inverse, algebraic, trig, and exponential.

Recall,

$$\text{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2,$$

where

$$\begin{aligned}
\mathbb{E}(X^2) &= \int_0^\infty x^2\exp(-x)dx \\
&= -x^2\exp(-x)|_0^\infty + \int_0^\infty 2x\exp(-x)dx \\
&= 2\int_0^\infty x\exp(-x) = 2 \text{ (from above)}, \\
&\therefore \text{Var}(X) = 2 - 1 = 1.
\end{aligned}$$

4.5

Find expectation, variance, and standard deviation of Y.

First, to get $\mathbb{E}(Y)$:

$$
\begin{aligned}
\mathbb{E}(Y) &= \int\limits_0^\infty \frac{y}{\lambda}\exp(\frac{-y}{\lambda})dy \\
&= \left[-\frac{\lambda y}{\lambda}\exp(\frac{-y}{\lambda})\right]_0^\infty + \int\limits_0^\infty \lambda\exp(\frac{-y}{\lambda})\frac{1}{\lambda}dy \\
&= \left[-y\exp(\frac{-y}{\lambda})\right]_0^\infty + \int\limits_0^\infty \exp(\frac{-y}{\lambda})dy \text{ (integrate by parts again)} \\
&= \left[-\lambda\exp(\frac{-y}{\lambda})\right]_0^\infty + \int_0^\infty \lambda\exp(\frac{-y}{\lambda})dy \\
&= \lambda.
\end{aligned}
$$

Then, for variance:

$$
\begin{aligned}
\mathrm{Var}(Y) = \mathbb{E}(Y-\lambda)^2 &= \int\limits_0^\infty (Y-\lambda)^2\frac{\exp(\frac{-y}{\lambda})}{\lambda}dy \\
&= \int\limits_0^\infty (y^2 - 2y\lambda + \lambda^2)\frac{\exp(\frac{-y}{\lambda})}{\lambda}dy \\
&= \left[-(y^2 - 2y\lambda + \lambda^2)\exp(\frac{-y}{\lambda})\right]_0^\infty + \int\limits_0^\infty \exp(\frac{-y}{\lambda})(2y + 2\lambda)dy, \\
&= \lambda^2.
\end{aligned}
$$

Thus, the standard deviation is λ.

There is a shorter method. Since we know $Y = \lambda X$, we can use the properties of the expectation and variance operators to quickly find $\mathbb{E}(Y)$ and $\mathrm{Var}(Y)$:

$$
\begin{aligned}
\mathbb{E}(X) = 1 &\implies \mathbb{E}(\lambda X) = \lambda, \\
\mathrm{Var}(X) = 1 &\implies \mathrm{Var}(\lambda X) = \lambda^2.
\end{aligned}
$$

4.6

Is it correct to say that $\mathbb{E}(1/X) = 1/\mathbb{E}(X)$?

Since $F(X)$ is a strictly convex function, and as $x \in \mathbb{R}^+$, we can apply Jensen's inequality (strictly) to get

$$
\mathbb{E}\left[\frac{1}{X}\right] > \frac{1}{\mathbb{E}(X)}.
$$

4.7

Use the change of variable formula to show that $F_X(x)$ has a uniform distribution. Hint: Find the inverse of the distribution by solving $z = F_X(x)$ for x.

We need to show that $F_Z(z) = 1$ using the change of variable formula. Suppose $z \sim U(0,1)$. Solve $z = 1 - \exp(-x)$ for x in terms of $z \in (0,1)$. First, we have

$$x = g^{-1}(z) = -\ln(1-z),$$

and

$$\begin{aligned} F_Z(z) &= F_X(g^{-1}(z)) \times |\frac{\partial}{\partial z} g^{-1}(z)| \\ &= \exp(\ln(1-z)) \frac{1}{1-z} \\ &= \frac{1-z}{1-z} = 1. \end{aligned}$$

5 Summations

The following formulas are used when discussing sample expectations and variances and regression on an intercept. Corresponding population versions exist. Prove the following.

5.1

$$\begin{aligned}
\sum_{i=1}^{n} (Y_i - \bar{Y}) &= \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} \bar{Y} \\
&= \sum_{i=1}^{n} Y_i - n\bar{Y} \\
&= \sum_{i=1}^{n} Y_i - \sum_{i=1}^{n} Y_i \\
&= 0.
\end{aligned}$$

5.2

$$\begin{aligned}
\sum_{i=1}^{n} (Y_i - \bar{Y})^2 &= \sum_{i=1}^{n} \left(Y_i^2 - Y_i\bar{Y} - \bar{Y}Y_i + \bar{Y}^2\right) \\
&= \sum_{i=1}^{n} Y_i \left(Y_i - \bar{Y}\right) - \sum_{i=1}^{n} \bar{Y} \left(Y_i - \bar{Y}\right) \\
&= \sum_{i=1}^{n} Y_i \left(Y_i - \bar{Y}\right) - \bar{Y} \sum_{i=1}^{n} \left(Y_i - \bar{Y}\right) \\
&= \sum_{i=1}^{n} Y_i \left(Y_i - \bar{Y}\right) - \bar{Y} \times 0 \\
&= \sum_{i=1}^{n} Y_i \left(Y_i - \bar{Y}\right).
\end{aligned}$$

From there, using the fact that $\sum_{i=1}^{n} Y_i = n\bar{Y}$:

$$\begin{aligned}
\sum_{i=1}^{n}(Y_i - \bar{Y})^2 &= \sum_{i=1}^{n} Y_i(Y_i - \bar{Y}) \\
&= \sum_{i=1}^{n} Y_i^2 - \sum_{i=1}^{n} Y_i\bar{Y} \\
&= \sum_{i=1}^{n} Y_i^2 - \bar{Y} \sum_{i=1}^{n} Y_i \\
&= \sum_{i=1}^{n} Y_i^2 - n\bar{Y}^2.
\end{aligned}$$

6 Sample variance and consistency

We show that the sample variance is consistent. This leads us to explore all the concepts in §3.1. Thus, the target is to prove that the sample variance is consistent for the population variance. Once again, pay attention to which assumptions are used where. Assume $X_1, ..., X_n$ *are IID with* $\mathbb{E}(X_i) = \mu$ *and* $Var(X_i) = \sigma^2$. *The sample variance is then*

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

6.1

Rewrite s^2 *as*

$$\frac{n}{n-1}\left[\sum_{i=1}^{n} X_i^2 - \left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2\right].$$

Begin with

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right],$$

using our result from Q4. Continuing on we have

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2\right],$$

since

$$\begin{aligned} n\bar{X}^2 = n\bar{X}\bar{X} &= \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)\left(\sum_{i=1}^{n} X_i\right) \\ &= \frac{n}{n^2}\left(\sum_{i=1}^{n} X_i\right)^2. \end{aligned}$$

So

$$\begin{aligned} s^2 &= \frac{1}{n-1}\left[\sum_{i=1}^{n} X_i^2 - \frac{n}{n^2}\left(\sum_{i=1}^{n} X_i\right)^2\right] \\ &= \frac{n}{n-1}\left[\frac{1}{n}\sum_{i=1}^{n} X_i^2 - \frac{1}{n^2}\left(\sum_{i=1}^{n} X_i\right)^2\right]. \end{aligned}$$

6.2

Argue that $n/(n-1) \to 1$.

Let $f(n) = \frac{n}{n-1} = \frac{u(n)}{v(n)}$, and take the limit as $n \to \infty$. We can then apply L'Hopital's rule to assess

$$\frac{u'(n)}{v'(n)} = 1,$$

thus completing the argument.

6.3

Argue that $\mathbb{E}(X_i^2) = \sigma^2 + \mu^2$.

Begin with

$$\begin{aligned} \mathbb{E}(X_i^2) &= \sigma^2 + \mu^2 = \text{Var}(X_i) + \mathbb{E}(X_i)^2 \\ \implies \text{Var}(X_i) &= \mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2, \end{aligned}$$

which we can factorise out to get:

$$\begin{aligned} \text{Var}(X_i) &= \mathbb{E}(X_i^2) - 2\mathbb{E}(X_i)^2 + \mathbb{E}(X_i)^2 \\ &= \mathbb{E}[X_i^2 - 2X_i\mathbb{E}(X_i) + \mathbb{E}(X_i)^2] \\ &= \mathbb{E}(X_i - \mathbb{E}(X_i))^2 \\ &= \text{Var}(X_i). \end{aligned}$$

6.4

Use the LLN on $n^{-1}\sum_{i=1}^{n} X_i$. *Carefully check assumptions.*

Define $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, then $\forall \epsilon > 0$:

$$\lim_{n\to\infty} P(|\bar{X} - \mu| < \epsilon) = 1,$$

in other words, the sample mean will tend to the population mean for a sufficiently large n. The proof comes from applying Chebychev's Inequality:

$$P(|\bar{X} - \mu| \geq \epsilon) = P((\bar{X} - \mu)^2 \geq \epsilon^2) \leq \frac{\mathbb{E}(\bar{X} - \mu)^2}{\epsilon^2} = \frac{\text{Var}(\bar{X})}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2},$$

Hence

$$P(|\bar{X} - \mu| < \epsilon) = 1 - P(|\bar{X} - \mu| \geq \epsilon) \geq 1 - \frac{\sigma^2}{n\epsilon^2} \to 1.$$

Thus,

$$\frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X} \xrightarrow{p} \mu.$$

6.5

Use LLN on $n^{-1}\sum_{i=1}^{n} X_i^2$. Carefully check assumptions.

Using our previous results,

$$\mathbb{E}(X_i^2) = \sigma^2 + \mu^2,$$

thus

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{p} \sigma^2 + \mu^2.$$

6.6

Argue that $s^2 \xrightarrow{p} \sigma^2$.

The continuous mapping theorem (CMT) (for vectors) shows that

$$s^2 = \underbrace{\frac{n}{n-1}}_{\to 1}\left[\underbrace{\frac{1}{n}\sum_{i=1}^{n} X_i^2}_{\xrightarrow{p}\sigma^2+\mu^2} - \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2}_{\xrightarrow{p}\mu^2}\right]$$

$$\therefore s^2 \xrightarrow{p} \sigma^2.$$

7 The t-ratio

We show that the t-statistic is asymptotically normal. Assume $X_1, ..., X_n$ are IID with $\mathbb{E}(X_i) = \mu$ and $Var(X_i) = \sigma^2$. Let $\bar{X} = n^{-1}\sum_{i=1}^n X_i$ and define $s^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \bar{X})^2$. We will show

$$z_n = \frac{\sqrt{n}(\bar{X} - \mu)}{s} \xrightarrow{d} N(0, 1).$$

7.1

Use the CLT to show that $\sqrt{n}(\bar{X} - \mu)/\sigma \xrightarrow{d} N(0, 1)$.

Since the X_i's are IID, we know $\mathbb{E}(X_i) = \mu$ and $\text{Var}(X_i) = \sigma^2$. Since $\bar{X} = n^{-1}\sum_{i=1}^n X_i$, $\bar{X}$ converges to μ. Furthermore, the standard deviation of $\bar{X}$ is $\sigma/\sqrt{n}$, so the distribution is collapsing around μ. If we don't want the distribution to collapse around μ, then we standardise $\bar{X}$ by multiplying the numerator and denominator by $\sqrt{n}$:

$$\bar{X}^* = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \xrightarrow{d} N(0, 1).$$

7.2

Recall that $s^2 \xrightarrow{p} \sigma^2$. Use the CMT to show that $z_n \xrightarrow{d} N(0, 1)$.

The limit of $\text{Var}(s^2) = 0$ as $n \to \infty$, and as previously discussed, $s^2 \xrightarrow{p} \sigma^2$, and $\sigma/s \xrightarrow{p} 1$. Hence:

$$\frac{\sqrt{n}(\bar{X} - \mu)}{s} = \frac{\sigma}{s}\frac{\sqrt{n}(\bar{X} - \mu)}{\sigma} \xrightarrow{d} N(0, 1)$$

via Slutsky's Theorem:

$$\frac{1}{\underbrace{s}_{=\sigma + o_p(1)}} \underbrace{\sqrt{n}\left(n^{-1}\sum_{i=1}^n (X_i - \mu)\right)}_{\xrightarrow{d} N(0,\sigma^2)} \xrightarrow{d} \frac{N(0, \sigma^2)}{\sigma} = N(0, 1).$$

We should note our key CMT results: If $X_n \xrightarrow{p} \mu$ and $Y_n \xrightarrow{d} F_Y(y)$, then $X_n \triangle Y_n \xrightarrow{d} \mu \triangle F_Y(y)$, where $\triangle$ denotes any operator $+ - \times \div$.