All of Statistics

My proposed solutions to the book:

All of Statistics - A Concise Course in Statistical Inference.
I really needed a proper refresher on statistics.

Let me know if you find any mistakes! I am sure there are plenty. :(

Contents

1 Chapter 5 - Convergence of Random Variables

Definition 5.1 Types of convergence
Let $X_1, X_2, \ldots$ be a sequence of random variables and let X be another random variable. Let F_n denote the CDF of X_n and let F denote the CDF of X.

1. X_n converges to X in *probability*, $X_n \xrightarrow{P} X$, if for all $\epsilon > 0$,

$$\mathbb{P}(|X_n - X| > \epsilon) \to 0$$

as $n \to \infty$.

2. X_n converges to X in *distribution*, $X_n \rightsquigarrow X$, if

$$\lim_{n\to\infty} F_n(t) = F(t)$$

at all t for which F is continuous.

Definition 5.2
X_n converges to X in *quadratic mean* (L_2), written $X_n \xrightarrow{qm} X$, if

$$\mathbb{E}[(X_n - X)^2] \to 0$$

as $n \to \infty$.

Relationship in convergence types.

$$\xrightarrow{qm} \implies \xrightarrow{P} \implies \rightsquigarrow$$

Exercises

5.1
Let $X_1, \ldots, X_n$ be IID with finite mean and variance μ and σ^2. Let $\overline{X}$ be the sample mean and S_n^2 be the sample variance.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n}, \qquad S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$$

(a) Showing that $\mathbb{E}[S_n^2] = \sigma^2$ was done in exercise 3.8.

(b) Showing that $S_n^2 \xrightarrow{P} \sigma^2$. Directly, this means that for any $\epsilon > 0$,

$$\mathbb{P}(|S_n^2 - \sigma^2| > \epsilon) = 0 \text{ as } n \to \infty.$$

But we can do it in a different way. Following the hint, we want to rewrite S_n^2 so we can apply the law of large numbers.

Doing the calculations on the next page.

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

Distributing the sum.

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

WLLN on $\overline{X}^2$ with $g(x) = x^2$.

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

The last rewrite is a standard identity. Now, we define $Y_i = X_i - \mu$. Then $\overline{Y} \to \mathbb{E}[(X_i - \mu)]$ by the WLLN. By Theorem 5.5(f), we can apply $g(x) = x^2$ and get $\overline{Y}^2 \to \mathbb{E}[(X_i - \mu)^2] = \mathrm{Var}(X_i) = \sigma^2$. Finally, we use the hint with $c_n = n/(n-1)$ which obviously tends to 1, so we can apply Theorem 5.5(d) and hence we have proved that $S_n^2 \xrightarrow{P} \sigma^2$. (The hint says using 5.5(e), but that is convergence with distribution, which I don't think is correct). (Update: this is a misprint which is noted in the errata2.pdf for the book).

5.2
Let $X_1, X_2 \ldots$ be a sequence of random variables. Show that $X_n \xrightarrow{qm} b$ if and only if

$$
\lim_{n\to\infty}\mathbb{E}[X_n] = b, \quad \text{and} \quad \lim_{n\to\infty}\mathrm{Var}[X_n] = 0.
$$

Proof.
$\Rightarrow$) We assume $X_n \xrightarrow{qm} b$, which means

$$
\mathbb{E}[(X_n - b)^2] \to 0, \quad n \to \infty.
$$

We can rewrite the expression:

$$
\begin{aligned}
\mathbb{E}[(X_n - b)^2] &= \mathbb{E}[X_n^2 - 2bX_n + b^2] \\
&= \mathbb{E}[X_n^2] - 2b\mathbb{E}[X_n] + b^2 \\
&= \mathbb{E}[X_n^2] - \mathbb{E}[X_n]^2 + \mathbb{E}[X_n]^2 - 2b\mathbb{E}[X_n] + b^2 \\
&= \mathrm{Var}(X_n) + (\mathbb{E}[X_n] - b)^2
\end{aligned}
$$

By our assumption $\mathbb{E}[(X_n - b)^2] \to 0$ as $n \to \infty$, and so $\text{Var}(X_n) + (\mathbb{E}[X_n] - b)^2 \to 0$ as $n \to \infty$. Since $\text{Var}(X_n) \geqslant 0$ and $(\mathbb{E}[X_n] - b)^2 \geqslant 0$, then we can conclude that

$$\text{Var}(X_n) \to 0, \qquad (\mathbb{E}[X_n] - b)^2 \to 0 \quad \Longrightarrow \quad \mathbb{E}[X_n] \to b$$

as $n \to \infty$.

$\Leftarrow$) We assume

$$\lim_{n\to\infty} \mathbb{E}[X_n] = b, \quad \text{and} \quad \lim_{n\to\infty} \text{Var}[X_n] = 0.$$

Then $\lim_{n\to\infty} \text{Var}(X_n) + (\mathbb{E}[X_n] - b)^2 = 0$. By reversing the calculations from the first part:

$$\lim_{n\to\infty} \text{Var}(X_n) + (\mathbb{E}[X_n] - b)^2 = 0 \quad \Longrightarrow \quad \lim_{n\to\infty} \mathbb{E}[(X_n - b)^2] = 0 \quad \Longrightarrow \quad X_n \xrightarrow{qm} b.$$

By showing implication both ways, the result is proved. $\square$

5.3

Let $X_1, \ldots, X_n$ be IID and let $\mu = \mathbb{E}[X_i]$ with finite variance. Show that $\overline{X} \xrightarrow{qm} \mu$.

Proof. Define:

$$\overline{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i.$$

By the WLLN, $\overline{X}_n \xrightarrow{P} \mu$, which we can also express as

$$\lim_{n\to\infty} \mathbb{E}[\overline{X}_n] = \mu.$$

The variance is finite, so $\text{Var}(X_i) = \sigma^2 < \infty$. This means that for the sample variance:

$$\lim_{n\to\infty} \text{Var}(\overline{X}) = \lim_{n\to\infty} \frac{\sigma^2}{n} = 0.$$

With this, we can simply apply the result from exercise 5.2 which shows that $\mathbb{E}[(\overline{X} - \mu)^2] \to 0$ as $n \to \infty$ which proves $\overline{X} \xrightarrow{qm} \mu$. $\square$

5.4

Let $X_1, X_2, \ldots$ be a sequence of random variables such that

$$\mathbb{P}\left(X_n = \frac{1}{n}\right) = 1 - \frac{1}{n^2}, \quad \text{and} \quad \mathbb{P}(X_n = n) = \frac{1}{n^2}.$$

Just noting the probabilities for each outcome as n grows.

	$n = 1$	$n = 2$	$n = 3$	$n = 4$	$\ldots$
$p(x) = \mathbb{P}(X_n = \frac{1}{n})$	$p(1) = 0$	$p(\frac{1}{2}) = \frac{3}{4}$	$p(\frac{1}{3}) = \frac{8}{9}$	$p(\frac{1}{4}) = \frac{15}{16}$	$\ldots$
$p(x) = \mathbb{P}(X_n = n)$	$p(1) = 1$	$p(2) = \frac{1}{4}$	$p(3) = \frac{1}{9}$	$p(4) = \frac{1}{16}$	$\ldots$

As n becomes large, the probability that we get n and not $1/n$ becomes very small. But as n becomes large, it will also yield extreme outliers in the sequence with a non-zero probability.

Starting by calculating the mean, second moment, and variance.

$$\begin{aligned}\mathbb{E}[X_n] &= \left(\frac{1}{n}\right)\left(1-\frac{1}{n^2}\right)+(n)\left(\frac{1}{n^2}\right)\\ &= \frac{1}{n}-\frac{1}{n^3}+\frac{1}{n}\\ &= \frac{2}{n}-\frac{1}{n^3}\end{aligned}$$

$$\begin{aligned}\mathbb{E}[X_n^2] &= \left(\frac{1}{n^2}\right)\left(1-\frac{1}{n^2}\right)+(n^2)\left(\frac{1}{n^2}\right)\\ &= \frac{1}{n^2}-\frac{1}{n^4}+1\end{aligned}$$

$$\begin{aligned}\mathrm{Var}(X_n) &= \mathbb{E}[X_n^2]-\mathbb{E}[X_n]^2\\ &= \frac{1}{n^2}-\frac{1}{n^4}+1-\left(\frac{2}{n}-\frac{1}{n^3}\right)^2\\ &= \frac{1}{n^2}-\frac{1}{n^4}+1-\left(\frac{4}{n^2}-\frac{4}{n^4}+\frac{1}{n^6}\right)\\ &= 1-\frac{3}{n^2}+\frac{3}{n^4}-\frac{1}{n^6}\end{aligned}$$

Results verified by simulation. (See code in `5.4.R`). E.g. for $n = 200$ and 10M simulations:

```
# Simulated vs. Theoretical
> mean(Xn)
[1] 0.009879878
> 2/n - 1/n^3
[1] 0.009999875
> var(Xn)
[1] 0.9759275
> 1 - 3/n^2 + 3/n^4 - 1/n^6
[1] 0.999925
```

R

From these expressions we can see that $\lim_{n\to\infty}\mathbb{E}[X_n]=0$, but $\lim_{n\to\infty}\mathrm{Var}(X_n)=1$. Since the variance does not become 0, we know by the result in exercise 5.2 that this does NOT converge in quadratic mean.

Checking if X_n converges in probability. If we fix some $\epsilon > 0$, we can apply the Chebyshev inequality. Set $\mu = \mathbb{E}[X_n]$ and $\sigma^2 = \mathrm{Var}(X_n)$:

$$\mathbb{P}(|X_n-\mu|>\epsilon) \leqslant \frac{\sigma^2}{\epsilon^2} = \frac{1-\frac{3}{n^2}+\frac{3}{n^4}-\frac{1}{n^6}}{\epsilon^2}$$

By taking the limit $n\to\infty$ on both sides, we get:

$$\lim_{n\to\infty}\mathbb{P}(|X_n-\mu|>\epsilon) \leqslant \lim_{n\to\infty}\frac{1-\frac{3}{n^2}+\frac{3}{n^4}-\frac{1}{n^6}}{\epsilon^2} = \frac{1}{\epsilon^2}.$$

Since this does not tend to 0, we do NOT have convergence in probability.

5.5
Let $X_1, \dots, X_n \sim \text{Bernoulli}(p)$. Prove that

$$\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{P} p, \quad \text{and} \quad \frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{qm} p.$$

Proof. Recalling how to calculate the expectation and second moment for a Bernoulli(p) variable:

$$\mathbb{E}[X] = (1)p + (0)(p-1) = p, \qquad \mathbb{E}[X^2] = (1)^2 p + (0)^2(p-1) = p,$$

With these, we can calculate the variance.

$$\text{Var}(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = p - p^2 = p(1-p).$$

We define $Y_i := X_i^2$ and exploit the simplicity of the Bernoulli distribution. In this case:

$$\mathbb{E}[Y] = \mathbb{E}[X^2] = (1)^2 p + (0)^2(p-1) = p, \qquad \mathbb{E}[Y^2] = \mathbb{E}[X^4] = (1)^4 p + (0)^4(p-1) = p,$$

With these, we can calculate the variance.

$$\text{Var}(Y) = \mathbb{E}[Y^2] - \mathbb{E}[Y]^2 = p - p^2 = p(1-p).$$

So we have $\mu = \mathbb{E}[Y] = p$ and $\sigma^2 = p(1-p)$. We can now define the sample mean and variance:

$$\mathbb{E}[\overline{Y}] = p, \qquad \text{Var}(\overline{Y}) = \frac{\sigma^2}{n} = \frac{p(1-p).}{n}.$$

Since the variance tends to 0 as $n \to \infty$ we can use the results in exercise 5.3 to conclude that $\overline{Y} \xrightarrow{qm} p$ and by how we defined Y, $\frac{1}{n}\sum_{i=1}^{n} X_i^2 \xrightarrow{qm} p$. Since we have convergence in quadratic mean, it follows that we also have convergence in probability. □

5.6
The height of men has mean 68 inches and standard deviation 2.6 inches. We have $n = 100$. Finding the approximate probability that the average height in the sample will be at least 68 inches.

We can approximate this probability with the CLT (central limit theorem). We want to find the probability that the sample height $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ is at least as big as the population mean: $\mathbb{P}(\overline{X} > \mu)$.

$$\mathbb{P}(\overline{X} > \mu) = 1 - \mathbb{P}(\overline{X} \leqslant \mu)$$

By the central limit theorem, using that $n = 100$, $\mu = 68$ and $\sigma = 2.6$:

$$\mathbb{P}(\overline{X} \leqslant 68) = \mathbb{P}(\overline{X} - 68 \leqslant 0) = \mathbb{P}\left(\frac{10(\overline{X} - 68)}{2.6} \leqslant 0\right) \approx \mathbb{P}(Z \leqslant 0) = 0.5$$

(Since the standard normal distribution is symmetric and centered at 0). So we get:

$$\mathbb{P}(\overline{X} > 68) = 1 - \mathbb{P}(\overline{X} \leqslant 68) = 1 - 0.5 = 0.5.$$

5.7
Let $\lambda_n = \frac{1}{n}$ for all n and let $X_n \sim \text{Poisson}(\lambda_n)$.

(a) Showing that $X_n \xrightarrow{P} 0$. By properties of the Poisson distribution, we have

$$\mu = \mathbb{E}[X_n] = \frac{1}{n}, \qquad \sigma^2 = \text{Var}(X_n) = \frac{1}{n}.$$

By fixing some $\epsilon > 0$ and applying Chebyshev's inequality, we get:

$$\mathbb{P}(|X_n - \frac{1}{n}| > \epsilon) = \mathbb{P}(|X_n - \mu| > \epsilon) \leqslant \frac{\sigma^2}{\epsilon^2} = \frac{1}{n\epsilon^2}$$

By taking the limit on both sides:

$$\lim_{n\to\infty} \mathbb{P}(|X_n - \frac{1}{n}| > \epsilon) =\leqslant \lim_{n\to\infty} \frac{1}{n\epsilon^2} = 0,$$

and since $\mu = 1/n \to 0$ we have shown that $X_n \xrightarrow{P} 0$.

(b) We define $Y_n = nX_n$ and will show that $Y_n \xrightarrow{P} 0$. Finding the mean and variance:

$$\mathbb{E}[Y_n] = \mathbb{E}[nX_n] = n\mathbb{E}[X_n] = n\left(\frac{1}{n}\right) = 1$$

$$\text{Var}(Y_n) = \text{Var}(nX_n) = n^2\text{Var}(X_n) = n^2\left(\frac{1}{n}\right) = n$$

From the Chebyshev inequality we can only really conclude that Y_n does NOT converge to 1, but we can't use it for determining the asymptotic behavior of Y_n at 0. Instead we will use Theorem 5.5(f):$X_n \xrightarrow{P} 0$ then $g(X_n) \xrightarrow{P} g(X)$. Here, Y_n is a function of X_n, defined as: $Y_n = nX_n$ where $g(x) = nx$, which is a continuous function. In this case we get that $X_n \xrightarrow{P} 0$ implies $nX_n \xrightarrow{P} n \cdot 0$ so $Y_n \xrightarrow{P} 0$.

5.8
A program has $n = 100$ pages of code. Let $X_i \sim \text{Poisson}(1)$ be iid and denote the number of errors on page i. Let $Y = \sum_{i=1}^{n} X_i$ denote the total number of errors. Use the CLT to approximate $\mathbb{P}(Y < 90)$.

The mean and variance are $\mu = \mathbb{E}[X_i] = 1$ and $\sigma^2 = \text{Var}(X_i) = 1$. An important observation: the CLT applies to sample means, but in this case we are simply summing up 100 independent Poisson variables, and so $Y \sim \text{Poisson}(100)$, i.e. $\mathbb{E}[Y] = 100$ and $\text{Var}(Y) = 100$.

Let us define $W = \frac{1}{n}Y$, which means:

$$W = \frac{1}{n}Y = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

then by the CLT, $W \sim N(1, 1/100)$, so $\mathbb{E}[W] = 1$ and $\text{Var}(W) = 1/100$.

By going the other way, we can find a normal approximation for Y by using that $Y = nW$ where $n = 100$.:

$$\mathbb{E}[Y] = \mathbb{E}[nW] = n\mathbb{E}[W] = n(1) = 100,$$

$$\mathrm{Var}(Y) = \mathrm{Var}(nW) = n^2\mathrm{Var}(W) = (100)^2\left(\frac{1}{100}\right) = 100$$

The standard deviation is: $\sqrt{\mathrm{Var}(Y)} = 10$. Approximating the probability.

$$\mathbb{P}(Y < 90) = \mathbb{P}\left(\frac{Y - 100}{10} < -9\right) \approx \mathbb{P}(Z \leqslant -9) \approx 0$$

Note: this exercise demonstrates that the CLT is just an approximation, and under certain conditions it's a very poor approximation. In `5.8.R` a simulation repeating the conditions were done one million times, and numerically, the probability that $\mathbb{P}(X < 90)$ turns out to be about 0.1467.

```
> # Approximating answer numerically
> length(Y)
[1] 1000000
> sum(Y < 90)/length(Y)
[1] 0.146773
```

R

5.9

Suppose that $\mathbb{P}(X = 1) = \mathbb{P}(X = -1) = 1/2$ and define:

$$X_n = \begin{cases} X & \text{with probability} \quad 1 - \frac{1}{n} \\ e^n & \text{with probability} \quad \frac{1}{n} \end{cases}$$