Introduction to Probability & Statistics

Assignment 5, 2024/25

Practice Questions

PQ1. You perform 28 independent experiments measuring a random variable X which you know has mean 457 and variance 676. Use Chebychev's inequality to give a lower bound on the probability that the mean of your measurements is between 433 and 481.

> Answer
>
> First we note that
> $$\mathbb{E}\left[\bar{X}_{28}\right] = \mathbb{E}\left[X\right] = 457$$
> and
> $$\mathsf{Var}\left(\bar{X}_{28}\right) = \frac{\mathsf{Var}\left(X\right)}{28} = \frac{676}{28}.$$
> Also we rewrite
> $$\mathbb{P}\left(433 < \bar{X}_{28} < 481\right) = \mathbb{P}\left(|\bar{X}_{28} - 457| < 24\right) = 1 - \mathbb{P}\left(|\bar{X}_{28} - \mathbb{E}\left[\bar{X}_{28}\right]| \geq 24\right).$$
> Chebychev's inequality tells us that
> $$\mathbb{P}\left(|\bar{X}_{28} - \mathbb{E}\left[\bar{X}_{28}\right]| \geq 24\right) \leq \frac{\mathsf{Var}\left(\bar{X}_{28}\right)}{(24)^2} = \frac{169}{4032}.$$
> Combining these two equations gives
> $$\mathbb{P}\left(433 < \bar{X}_{28} < 481\right) \geq 1 - \frac{169}{4032} = \frac{3863}{4032} \approx 0.96.$$

PQ2. Let $X_1, \ldots, X_n$ be an i.i.d. sample from the $\mathsf{N}(\mu, \sigma^2)$ distribution.

a) Write down a $100(1-\alpha)\%$ confidence interval estimator for μ in the case where σ^2 is known. Hence or otherwise, find a 95% confidence interval for μ in the case where $\sigma^2 = 9$ and a random sample of size 16 has been taken with values $x_1, \ldots, x_{16}$ and it has been found that
$$\sum_{i=1}^{16} x_i = 50, \quad \sum_{i=1}^{16} (x_i - \bar{x}_{16})^2 = 115.$$

b) It is proposed that, from a second independent random sample of size 16 a 99% confidence interval for μ be constructed and that, from a third independent random sample of size 32, a 98% confidence interval for μ be constructed. State the probability that *neither* of these two confidence intervals will contain μ.

Answer

a) The confidence interval estimator is

$$\left(\bar{X}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{X}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right).$$

With $\sigma = 3$, $n = 16$, $\bar{x}_n = 50/16$ and $z_{0.025} = 1.96$ substituted into the above expression we get the confidence interval $(1.655, 4.595)$.

b) The event that the second sample leads to a confidence interval that does not contain μ and the event that the third sample leads to a confidence interval that does not contain μ are independent because the samples are independent. Thus the probability that both samples lead to a confidence interval that does not contain μ is equal to the product of the individual probabilities, $0.01 \cdot 0.02 = 0.0002$.

PQ3. Lengths of baguettes are assumed to follow a $\mathsf{N}(\mu, \sigma^2)$ distribution. Six baguettes were measured, giving the following lengths in cms: 66, 69, 62, 62, 64, 67.

a) Calculate unbiased estimates for μ and σ^2.

b) Calculate a 90% confidence interval for μ.

Answer

a) An unbiased estimate for μ is $\hat{\mu} = \bar{x} = 65cm$. An unbiased estimate for σ^2 is $\hat{\sigma^2} = s_n^2 = 8$.

b) A 90% confidence interval is given by

$$\begin{aligned}\left(\bar{x}_n - t_{n-1,\alpha/2}\frac{s_n}{\sqrt{n}}, \bar{x}_n + t_{n-1,\alpha/2}\frac{s_n}{\sqrt{n}}\right) &= \left(65 - t_{5,0.05}\sqrt{8/6}, 65 + t_{5,0.05}\sqrt{8/6}\right)\\ &= \left(65 - 2.02\sqrt{8/6}, 65 + 2.02\sqrt{8/6}\right)\\ &= (62.67, 67.33)\,.\end{aligned}$$

PQ4. Let $X_1, \ldots, X_n$ be an i.i.d. sample from a distribution with mean μ and variance σ^2. Let $\bar{X}_n$ denote the sample mean, and define

$$Z_n = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma}\,.$$

Check that $\mathbb{E}[Z_n] = 0$ and $\mathsf{Var}(Z_n) = 1$.

Answer

We know that $\mathbb{E}\left[\bar{X}_n\right] = \mu$ and $\mathsf{Var}\left(\bar{X}_n\right) = \sigma^2/n$. By linearity of expectation we have

$$\mathbb{E}[Z_n] = \frac{\sqrt{n}\mathbb{E}\left[\bar{X}_n - \mu\right]}{\sigma} = 0\,.$$

For the variance, we know that for any random variable Y, and constants $a, b \in \mathbb{R}$, $\mathsf{Var}(aY + b) = a^2\mathsf{Var}(Y)$. Thus

$$\mathsf{Var}(Z_n) = \mathsf{Var}\left(\sqrt{n}\bar{X}_n/\sigma - \sqrt{n}\mu/\sigma\right) = (\sqrt{n}/\sigma)^2\mathsf{Var}\left(\bar{X}_n\right) = 1\,.$$

PQ5. Recall that we say that T_m has a t-distribution with $m > 1$ degrees of freedom, and write $T_m \sim t(m)$, if it has density function given by

$$f(x) = k_m \left(1 + \frac{x^2}{m}\right)^{-\frac{m+1}{2}}, \qquad x \in \mathbb{R},$$

where k_m is a constant that ensures that the density integrates to 1. If $T_m \sim t(m)$, show that $\mathbb{E}[T_m] = 0$.
Hint: you shouldn't need to explicitly calculate any integrals here!

Answer

We note that $f(x)$ is an even function: $f(-x) = f(x)$. Thus

$$\begin{aligned}
\mathbb{E}[T_m] &= \int_{-\infty}^{\infty} x f(x) dx \\
&= \int_{-\infty}^{0} x f(x) dx + \int_{0}^{\infty} x f(x) dx \\
&= \int_{0}^{\infty} (-x) f(-x) dx + \int_{0}^{\infty} x f(x) dx \\
&= \int_{0}^{\infty} (-x) f(x) dx + \int_{0}^{\infty} x f(x) dx \\
&= \int_{0}^{\infty} ((-x) + x) f(x) dx \\
&= 0\,.
\end{aligned}$$

Assignment Questions – answers to be uploaded

AQ1. Let $X_1, \ldots, X_{16}$ be an i.i.d. sample from a $\mathsf{N}(3, 1)$ distribution, and let $S = X_1 + X_2 + \cdots + X_{16}$. Express $\mathbb{P}(S < 52)$ in terms of the distribution function Φ of the standard normal distribution.

Answer

We know that $\mathbb{E}[S] = 16\mathbb{E}[X] = 48$, and $\mathsf{Var}(S) = 16\mathsf{Var}(X) = 16$ (because the X_i are independent, and hence uncorrelated).
Furthermore, we know that the sum of independent normal distributions has a normal distribution. Thus

$$S \sim \mathsf{N}(48, 16)\,.$$

[3]
We can normalize to obtain a standard normal random variable by subtracting the mean and dividing by the standard deviation. Thus

$$\mathbb{P}(S < 52) = \mathbb{P}\left(\frac{S - 48}{\sqrt{16}} < \frac{52 - 48}{\sqrt{16}}\right) = \Phi(1)\,.$$

[2]

AQ2. Let $Y_1, Y_2, \ldots$ be an i.i.d. sequence of random variables, each with a Uniform$(0, 3)$ distribution. Define a new sequence of random variables $X_1, X_2, \ldots$ by

$$X_n = \frac{1}{n} \sum_{i=1}^{n} Y_i^2\,.$$

Using the Law of Large Numbers, determine the value of $a \in \mathbb{R}$ for which $\mathbb{P}(\lim_{n\to\infty} X_n = a) = 1$.

Answer

The random variable X_n is just the sample mean of the random variables $Y_1^2, \ldots, Y_n^2$. These are i.i.d. and clearly have finite mean and variance (since Y_i^2 can only take values in the finite set $[0, 9]$).
The (strong) law of large numbers says that, with probability one, X_n will converge to $\mathbb{E}\left[Y^2\right]$. **[3]**
Finally, we calculate

$$\mathbb{E}\left[Y^2\right] = \int_0^3 y^2 \frac{1}{3} dy = 3\,,$$

and so the required answer is $a = 3$. **[2]**

AQ3. Suppose the random variables X_1, X_2 and X_3 all have the same expectation μ. For what values of a and b is

$$M = -4(X_1 - 2) + 9(X_2 - 1) + aX_3 + b$$

an unbiased estimator for μ?

Answer

M is an unbiased estimator for μ if $\mathbb{E}\left[M\right] = \mu$ for any value of μ. **[1]** We find

$$\begin{aligned}
\mathbb{E}\left[M\right] &= \mathbb{E}\left[-4(X_1 - 2) + 9(X_2 - 1) + aX_3 + b\right] \\
&= -4(\mathbb{E}\left[X_1\right] - 2) + 9(\mathbb{E}\left[X_2\right] - 1) + a\mathbb{E}\left[X_3\right] + b \\
&= -4(\mu - 2) + 9(\mu - 1) + a\mu + b \\
&= (5 + a)\mu + (b - 1).
\end{aligned}$$

[2]
Thus $\mathbb{E}\left[\mu\right] = \mu$ if and only if $a = -4$ and $b = 1$. **[2]**

AQ4. From a dataset $x_1, \ldots, x_{10}$ it has been calculated that

$$\sum_{i=1}^{10} x_i = 491, \quad \sum_{i=1}^{10} (x_i - \bar{x}_{10})^2 = 41.$$

You model the dataset as a random sample from the normal distribution with mean μ and variance σ^2.

a) Assume that both μ and σ^2 are unknown. Determine a 95% confidence interval for the mean μ. You can use that $t_{9,0.025} \approx 2.26$.

b) Now assume that it is known that the variance is $\sigma^2 = 5$. Give a 95% confidence interval for the mean μ in this case. You can use that $z_{0.025} \approx 1.96$.

Answer

a) Since the variance is unknown, we use the interval

$$\left(\bar{x}_n - t_{n-1,\alpha/2}\frac{s_n}{\sqrt{n}}, \bar{x}_n + t_{n-1,\alpha/2}\frac{s_n}{\sqrt{n}}\right)$$

We first calculate $\bar{x}_n = 491/10 = 49.1$ and

$$s_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x}_n)^2 = \frac{1}{10-1}41 = \frac{41}{9}.$$

We're told that $t_{n-1,\alpha/2} = t_{9,0.025} = 2.26$ and

$$\left(\bar{x}_n - t_{n-1,\alpha/2}\frac{s_n}{\sqrt{n}}, \bar{x}_n + t_{n-1,\alpha/2}\frac{s_n}{\sqrt{n}}\right) \approx \left(49.1 - 2.26\frac{\sqrt{41}}{3\sqrt{10}}, 49.1 + 2.26\frac{\sqrt{41}}{3\sqrt{10}}\right)$$
$$\approx (47.57, 50.63).$$

[3]

b) When σ^2 is known, we use the confidence interval

$$\left(\bar{x}_n - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \bar{x}_n + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) \approx \left(49.1 - 1.96\frac{\sqrt{5}}{\sqrt{10}}, 49.1 + 1.96\frac{\sqrt{5}}{\sqrt{10}}\right)$$
$$\approx (47.71, 50.49).$$

[2]

Other Questions (for seminars / extra practice)

OQ1. If X has expectation μ and standard deviation σ, the ratio $r = |\mu|/\sigma$ is called the *measurement signal-to-noise-ratio* of X. If we define $D = |(X-\mu)/\mu|$ as the *relative deviation* of X from its mean μ, show that, for $\alpha > 0$,

$$\mathbb{P}(D < \alpha) \geq 1 - \frac{1}{r^2\alpha^2}.$$

Answer

$$\mathbb{P}(D < \alpha) = \mathbb{P}(|(X-\mu)/\mu| < \alpha) = \mathbb{P}(|X-\mu| < \alpha r\sigma)\ .$$

We now use Chebychev's inequality:

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

with $k = \alpha r$, giving

$$\mathbb{P}(|X-\mu| < \alpha r\sigma) \geq 1 - \frac{1}{r^2\alpha^2}.$$

OQ2. Let X be the number of 1s and Y be the number of 2s that occur in n rolls of a fair die. Use indicator random variables to compute $\mathrm{Cov}(X, Y)$ and $\rho(X, Y)$.
Hint: this is just like the smarties example covered in lectures.

Answer

We introduce the indicator random variables

$$X_i = \begin{cases} 1 & \text{roll } i \text{ lands on 1} \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad Y_i = \begin{cases} 1 & \text{roll } i \text{ lands on 2} \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$X = \sum_{i=1}^{n} X_i \quad \text{and} \quad Y = \sum_{i=1}^{n} Y_i.$$

We observe that $\mathbb{P}(X_i = 1) = \mathbb{P}(Y_i = 1) = 1/6$ and $\mathbb{P}(X_iY_j = 1) = 1/36$ if $i \neq j$ and $\mathbb{P}(X_iY_i = 1) = 0$. Thus $\mathbb{E}[X_i] = \mathbb{E}[Y_i] = 1/6$ and $\mathbb{E}[X_iY_j] = 1/36$ if $i \neq j$ and $\mathbb{E}[X_iY_i] = 0$.

We calculate the covariances:

$$\text{Cov}\left(X_i, Y_j\right) = \mathbb{E}\left[X_i Y_j\right] - \mathbb{E}\left[X_i\right]\mathbb{E}\left[Y_j\right] = \begin{cases} \frac{1}{36} - \frac{1}{6}\cdot\frac{1}{6} = 0 & \text{if } i \neq j \\ -\frac{1}{6}\cdot\frac{1}{6} = \frac{-1}{36} & \text{if } i = j. \end{cases}$$

Now, as we've seen with the smarties example in lectures,

$$\begin{aligned} \text{Cov}\left(X, Y\right) = \text{Cov}\left(\sum_{i=1}^{n} X_i, \sum_{j=1}^{n} Y_j\right) &= \sum_{i=1}^{n}\sum_{j=1}^{n} \text{Cov}\left(X_i, Y_j\right) \\ &= \sum_{i=1}^{n} \text{Cov}\left(X_i, Y_i\right) = \sum_{i=1}^{n} \frac{-1}{36} = -\frac{n}{36}. \end{aligned}$$

To calculate the variance of X we recall that the variance of a sum of independent random variables equals the sum of their variances. Thus

$$\text{Var}\left(X\right) = \text{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \text{Var}\left(X_i\right) = \sum_{i=1}^{n} \frac{5}{36} = \frac{5n}{36}.$$

By the same calculation $\text{Var}\left(Y\right)$ has the same value. (We could also have used the fact that X and Y each have a $\text{Bin}(n, 1/6)$ distribution.) Thus

$$\rho(X, Y) = \frac{\text{Cov}\left(X, Y\right)}{\sqrt{\text{Var}\left(X\right)\text{Var}\left(Y\right)}} = \frac{-n/36}{5n/36} = -\frac{1}{5}.$$

It makes sense that the correlation coefficient should be negative, because if in n dice rolls I get an unusually high number of 1s, then the chance of getting many 2s is smaller (because all those 1s are definitely not 2s).

OQ3. In the lecture I proved Chebychev's inequality for the case of a continuous random variable. Provide a similar proof for the case of a discrete random variable.

Answer

First we find a lower bound for the variance of X. Let's write $\mathbb{E}\left[X\right] = \mu$.

$$\begin{aligned} \text{Var}\left(X\right) = \mathbb{E}\left[(X-\mu)^2)\right] &= \sum_{x \in X(\Omega)} (x-\mu)^2 p_X(x) \\ &\geq \sum_{\substack{x \in X(\Omega) \\ |x-\mu| \geq a}} (x-\mu)^2 p_X(x) \\ &\geq \sum_{\substack{x \in X(\Omega) \\ |x-\mu| \geq a}} a^2 p_X(x) = a^2 P(|X-\mu| \geq a). \end{aligned}$$

Then we divide both sides of the inequality by a^2 to get

$$\mathbb{P}\left(\left|X - \mathbb{E}\left[X\right]\right| \geq a\right) \leq \frac{1}{a^2}\text{Var}\left(X\right).$$

OQ4. Assume that in Example 17.1 from the lectures (measuring a ball rolling down an inclined plane), we choose to stop the ball always after one time unit, so that

$$X_i = \frac{1}{2}a\left(1 + U_i\right)^2 + V_i,$$

where the independent errors are normally distributed with $U_i \sim \text{N}(0, \sigma_U^2)$, $V_i \sim \text{N}(0, \sigma_V^2)$. Assume the variances of the errors are known. Calculate the bias of the estimator $A = 2\bar{X}_n$ for the acceleration parameter a. Propose an unbiased estimator for a.

Answer

For the bias we calculate

$$\mathbb{E}[A] = \mathbb{E}\left[2\bar{X}_n\right] = 2\mathbb{E}\left[\bar{X}_n\right] = 2\mathbb{E}[X_i].$$

So we need the expectation of X_i:

$$\begin{aligned}\mathbb{E}[X_i] &= \mathbb{E}\left[\frac{1}{2}a(1+U_i)^2 + V_i\right] = \frac{1}{2}a\mathbb{E}\left[(1+U_i)^2\right] + \mathbb{E}[V_i] \\ &= \frac{1}{2}a\left(\text{Var}(1+U_i) + \mathbb{E}[1+U_i]^2\right) = \frac{1}{2}a\left(\text{Var}(U_i) + 1^2\right) \\ &= \frac{1}{2}a\left(\sigma_U^2 + 1\right).\end{aligned}$$

This gives

$$\mathbb{E}[A] = 2\mathbb{E}[X_i] = a(\sigma_U^2 + 1) \neq a.$$

So this estimator is not unbiased. Taking the average is going to consistently overestimate the value of a.
Luckily we can fix this by rescaling the estimator. The estimator

$$\tilde{A} = \frac{1}{\sigma_U^2 + 1}A = \frac{2}{\sigma_U^2 + 1}\bar{X}_n$$

is unbiased.

OQ5. Consider the following dataset of lifetimes of ball bearings in hours:

$$\begin{array}{cccccccccc} 6278 & 3113 & 5236 & 11584 & 12628 & 7725 & 8604 & 14266 & 6125 & 9350 \\ 3212 & 9003 & 3523 & 12888 & 9460 & 13431 & 17809 & 2812 & 11825 & 2398. \end{array}$$

Suppose that we are interested in estimating the minimum lifetime of this type of ball bearing. The dataset is modelled as a realization of a random sample $X_1, \ldots, X_n$. Each random variable X_i is represented as $X_i = \delta + Y_i$, where Y_i has an Exp(λ) distribution and $\delta > 0$ is an unknown parameter that is supposed to model the minimum lifetime. The objective is to construct an unbiased estimator for δ. It is known that

$$\mathbb{E}[M_n] = \delta + \frac{1}{n\lambda} \quad \text{and} \quad \mathbb{E}\left[\bar{X}_n\right] = \delta + \frac{1}{\lambda},$$

where $M_n = \min(X_1, \ldots, X_n)$ and $\bar{X}_n = (X_1 + \cdots + X_n)/n$.

a) Check whether

$$T = \frac{n}{n-1}\left(\bar{X}_n - M_n\right)$$

is an unbiased estimator for $1/\lambda$.

b) Construct an unbiased estimator D for δ.

c) Use the dataset to compute an estimate for the minimum lifetime δ.

Answer

a) By linearity of expectation

$$\mathbb{E}[T] = \frac{n}{n-1}\left(\mathbb{E}\left[\bar{X}_n\right] - \mathbb{E}[M_n]\right)$$
$$= \frac{n}{n-1}\left(\left(\delta + \frac{1}{\lambda}\right) - \left(\delta + \frac{1}{n\lambda}\right)\right)$$
$$= \frac{n}{n-1}\frac{n-1}{n\lambda} = \frac{1}{\lambda}.$$

This shows that T is an unbiased estimator for $1/\lambda$.

b) We will look for a linear combination of $\bar{X}_n$ and M_n of which the expectation is δ. From the expressions for $\mathbb{E}\left[\bar{X}_n\right]$ and $\mathbb{E}[M_n]$ we see that we can eliminate λ by subtracting $\mathbb{E}\left[\bar{X}_n\right]$ from $n\mathbb{E}[M_n]$. Therefore, first consider $nM_n - \bar{X}_n$, which has expectation

$$\mathbb{E}\left[nM_n - \bar{X}_n\right] = n\mathbb{E}[M_n] - \mathbb{E}\left[\bar{X}_n\right] = n\left(\delta + \frac{1}{n\lambda}\right) - \left(\delta + \frac{1}{\lambda}\right) = (n-1)\delta.$$

This means that
$$D = \frac{nM_n - \bar{X}_n}{n-1}$$
has expectation δ: $\mathbb{E}[D] = \mathbb{E}\left[nM_n - \bar{X}_n\right]/(n-1) = \delta$, so that D is an unbiased estimator for δ.

c) When we evaluate $\bar{X}_n$ and M_n on the dataset with $n = 20$ values, we find

$$\bar{x}_n = 8563.5, \qquad m_n = 2398.$$

Substituting these values into the estimator determined in part b) gives the following estimate for δ:
$$\frac{nm_n - \bar{x}_n}{n-1} = \frac{20 \cdot 2398 - 8563.5}{19} = 2073.5.$$

OQ6. Let M_n be the maximum of n independent Uniform$(0,1)$ random variables. Show that for any fixed $\varepsilon > 0$,
$$\lim_{n\to\infty} \mathbb{P}(|M_n - 1| > \varepsilon) = 0\,.$$

Answer

Let $X_1, \ldots, X_n$ denote the n independent Uniform$(0,1)$ random variables. M_n can only take values in the set $(0,1)$, so the event $\{|M_n - 1| > \varepsilon\} = \{M_n < 1 - \varepsilon\}$. Now note that for any $x \in (0,1)$, $\{M_n < x\}$ occurs if and only if $\{X_i < x\}$ for *all* of the n uniform random variables. Since the X_i are independent, we get

$$\mathbb{P}(M_n < 1 - \varepsilon) = \mathbb{P}(X < 1 - \varepsilon)^n = (1 - \varepsilon)^n\,,$$

where in the last line we have used the distribution function for a Uniform$(0,1)$ distribution to calculate $\mathbb{P}(X < 1 - \varepsilon)$. This tends to 0 as $n \to \infty$, as required.

OQ7. [Harder] (A more general law of large numbers, see Exercise 13.12 in the textbook).
Let $X_1, X_2, \ldots$ be a sequence of independent random variables with $\mathbb{E}[X_i] = \mu_i$ and Var$(X_i) = \sigma_i^2$ for $i = 1, 2, \ldots$. Let $\bar{X}_n = (X_1 + \cdots + X_n)/n$. Suppose that there exists an $M \in \mathbb{R}$ such that $0 < \sigma_i^2 \leq M$ for all i, and let a be an arbitrary positive number.

a) Apply Chebychev's inequality to show that

$$\mathbb{P}\left(\left|\bar{X}_n - \frac{1}{n}\sum_{i=1}^{n}\mu_i\right| > a\right) \le \frac{\mathrm{Var}\,(X_1) + \cdots + \mathrm{Var}\,(X_n)}{n^2a^2}.$$

b) Conclude from a) that

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\bar{X}_n - \frac{1}{n}\sum_{i=1}^{n}\mu_i\right| > a\right) = 0.$$

c) Check that the weak law of large numbers is a special case of this result.

Answer

a) First we calculate the expectation and variance of $\bar{X}_n$:

$$\begin{aligned}
\mathbb{E}\left[\bar{X}_n\right] &= \mathbb{E}\left[(X_1 + \cdots X_n)/n\right] \\
&= \left(\mathbb{E}\left[X_1\right] + \cdots + \mathbb{E}\left[X_n\right]\right)/n \text{ by linearity of expectation} \\
&= (\mu_1 + \cdots + \mu_n)/n = \frac{1}{n}\sum_{i=1}^{n}\mu_i, \\
\mathrm{Var}\left(\bar{X}_n\right) &= \mathrm{Var}\left((X_1 + \cdots + X_n)/n\right) \\
&= \mathrm{Var}\,(X_1/n) + \cdots + \mathrm{Var}\,(X_n/n) \text{ using independence of the } X_i \\
&= \frac{\mathrm{Var}\,(X_1) + \cdots + \mathrm{Var}\,(X_n)}{n^2}.
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}\left(\left|\bar{X}_n - \frac{1}{n}\sum_{i=1}^{n}\mu_i\right| > a\right) &= \mathbb{P}\left(\left|\bar{X}_n - \mathbb{E}\left[\bar{X}_n\right]\right| > a\right) \\
&\le \mathbb{P}\left(\left|\bar{X}_n - \mathbb{E}\left[\bar{X}_n\right]\right| \ge a\right) \\
&\le \frac{1}{a^2}\mathrm{Var}\left(\bar{X}_n\right) \text{ by Chebychev} \\
&= \frac{\mathrm{Var}\,(X_1) + \cdots + \mathrm{Var}\,(X_n)}{n^2a^2}.
\end{aligned}$$

b) On the right-hand side of the inequality from part a) we can use that $\mathrm{Var}\,(X_i) = \sigma_i^2 \le M$ for all i and hence

$$\frac{\mathrm{Var}\,(X_1) + \cdots + \mathrm{Var}\,(X_n)}{n^2a^2} \le \frac{nM}{n^2a^2} = \frac{M}{na^2}.$$

We now take the limit $n \to \infty$ on both sides of the inequality

$$\mathbb{P}\left(\left|\bar{X}_n - \frac{1}{n}\sum_{i=1}^{n}\mu_i\right| > a\right) \le \frac{M}{na^2}$$

and use that $\lim_{n\to\infty} M/(na^2) = 0$ to obtain

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\bar{X}_n - \frac{1}{n}\sum_{i=1}^{n}\mu_i\right| > a\right) \le 0.$$

But we also know that probabilities are always non-negative, and the limit of a sequence of non-negative numbers is also non-negative, so the limit must be zero.

c) The law of large numbers has the same assumptions as the statement in this question, with the additional requirement that all the X_i are identically distributed so that in particular they all have the same expectation $\mu_i = \mu$. Thus the average of all the μ_i is also μ and the statement from part b) becomes

$$\lim_{n\to\infty} \mathbb{P}\left(\left|\bar{X}_n - \mu\right| > a\right) = 0$$

for any $a > 0$. This is the statement of the weak law of large numbers.