Homework - Practice - Week 1

Probabilistic Methods

May 2025

Contents

Exercise 1: The Normal Distribution

(a) Sampling, histograms and QQ-plots

Figure 1 shows the two histograms for samples of size 100 and 100 000 drawn from $\mathcal{N}(0, 1)$, while Fig. 2 displays the corresponding QQ-plots.

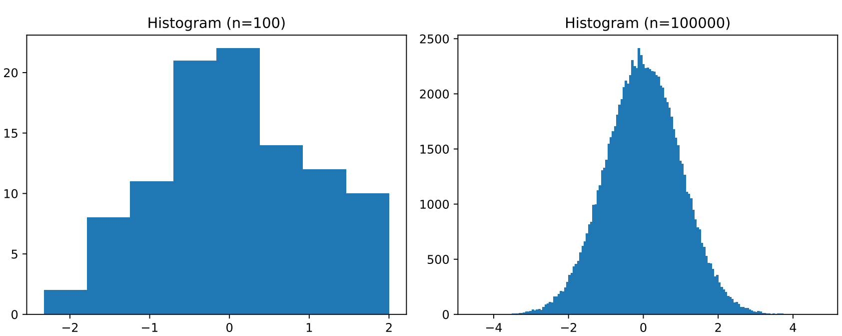

Figure 1: Side-by-side histograms for $n = 100$ (left) and $n = 100\,000$ (right).

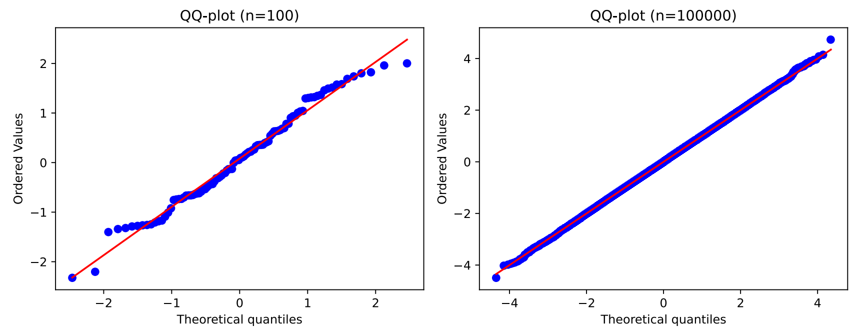

Figure 2: QQ-plots for $n = 100$ (left) and $n = 100\,000$ (right).

Estimated means and standard deviations:

Parameter	n=100	n=100 000
Mean $\hat{\mu}$	0.017	-0.004
SD $\hat{\sigma}$	1.080	0.996

(b) Exact probabilities for $Z \sim \mathcal{N}(0, 1)$

$$P(Z < 2) = 0.977, \qquad P(Z > -0.5) = 0.691, \qquad P(-1 < Z < 2) = 0.819.$$

(c) Monte-Carlo estimates

Event	Exact	n=100	n=100 000
$Z < 2$	0.977	0.950	0.978
$Z > -0.5$	0.691	0.670	0.691
$-1 < Z < 2$	0.819	0.770	0.819

(d) The distribution $\mathcal{N}(3, 2^2)$

Estimated parameters from two new samples:

Parameter	n=100	n=100 000
Mean $\hat{\mu}$	2.961	3.001
SD $\hat{\sigma}$	1.710	2.002

Exact probabilities for $X \sim \mathcal{N}(3, 4)$:

$$P(X < 2) = 0.309, \qquad P(X > -0.5) = 0.960, \qquad P(-1 < X < 2) = 0.286.$$

The 95$^{\text{th}}$ percentile equals 6.29.

(e) Generating $\mathcal{N}(-10, 5^2)$

A sample of size 1000 produced via $X = -10+5Z$ (with $Z \sim N(0, 1)$) yields $\bar{x} \approx -9.89$ and $s \approx 5.08$, as expected.

Exercise 2: Other Distributions

For each distribution the histogram, box-plot and QQ-plot are stacked into one PDF; see Figures 3–6.

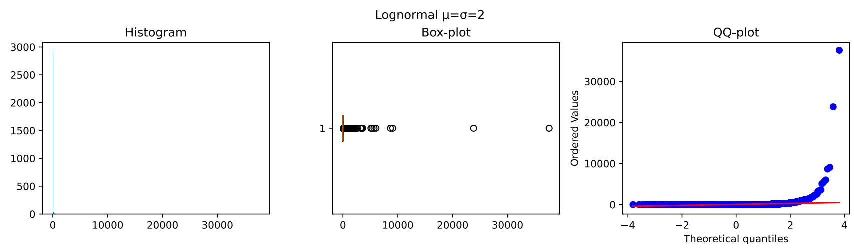

Figure 3: Log-normal sample ($n = 10\,000$, $\mu = \sigma = 2$).

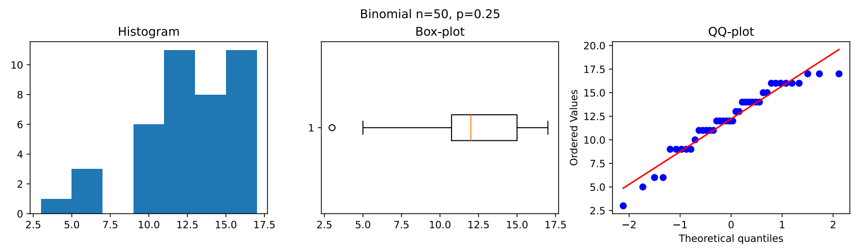

Figure 4: Binomial sample ($n = 40$, $n_{\text{trial}} = 50$, $p = 0.25$).

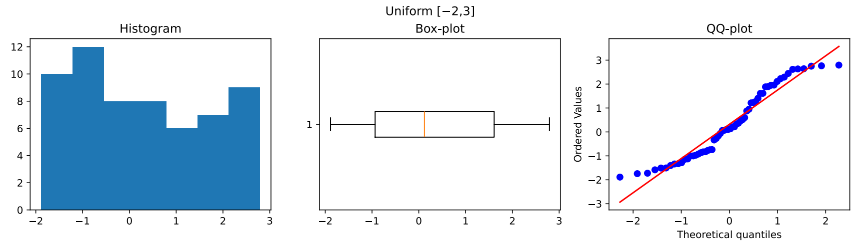

Figure 5: Uniform sample ($n = 60$, range $[-2, 3]$).

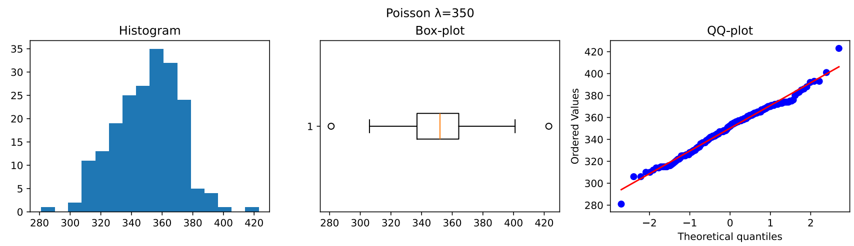

Figure 6: Poisson sample ($n = 200$, $\lambda = 350$).

Exercise 3: Teen Birth and Mortality Data

(a) Numerical summaries

Statistic	Teen birth	Mortality
Mean	12.43	10.3
SD	3.293	1.35
Variance	10.844	1.822
Range	[7.3, 20.5]	[8.4, 13.3]

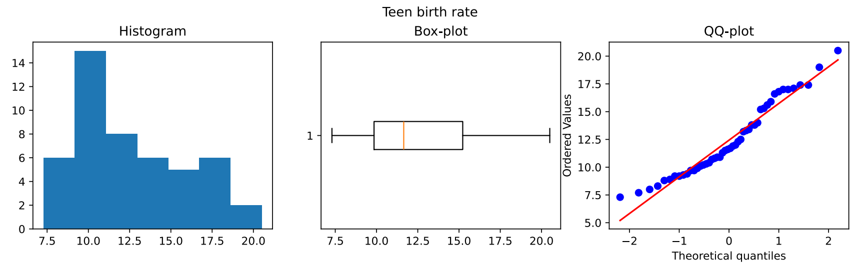

Figure 7: Distribution of teen birth rates.

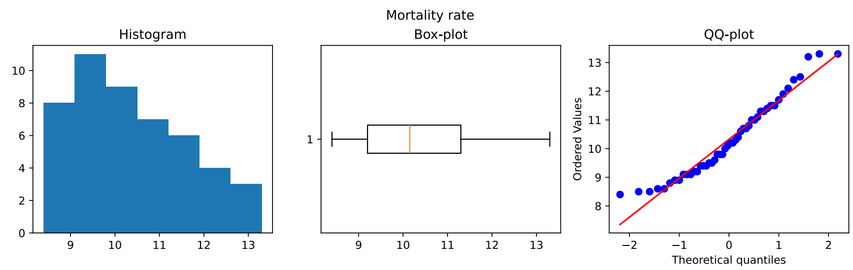

Figure 8: Distribution of mortality rates.

(b) Correlation

The empirical correlation between the two variables is $\rho \approx 0.549$.

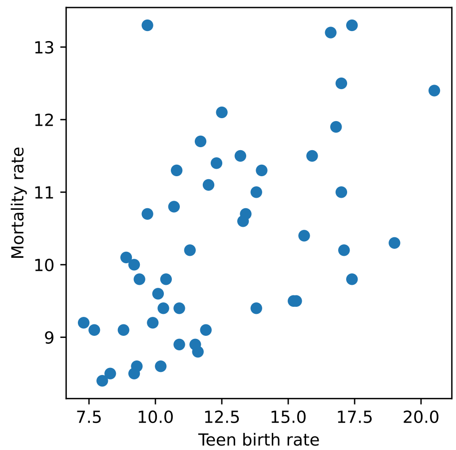

Figure 9: Scatter plot of teen birth versus mortality rates.

Exercise 4: Simulating p −values for a t-test

(a)–(b) Null hypothesis true ($\mu = \nu = 180$)

SD	$\Pr(p < 0.05)$	$\Pr(p < 0.10)$
10	0.046	0.103
1	0.048	0.099

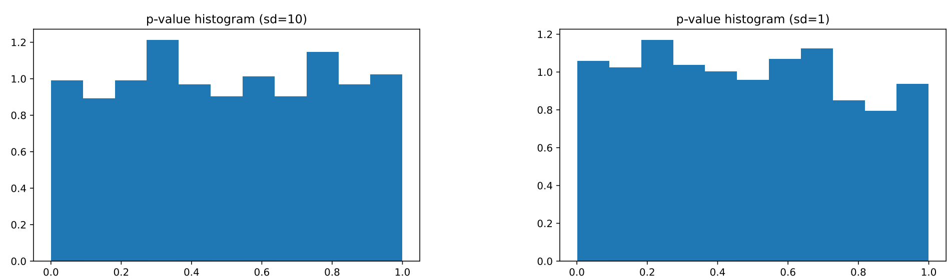

Figure 10: Histograms of p-values when H_0 is true (left: $\sigma = 10$, right: $\sigma = 1$).

(c) Null hypothesis false ($\mu = 180$, $\nu = 175$, $\sigma = 6$)

$$\Pr(p < 0.05) = 0.881, \qquad \Pr(p < 0.10) = 0.929.$$

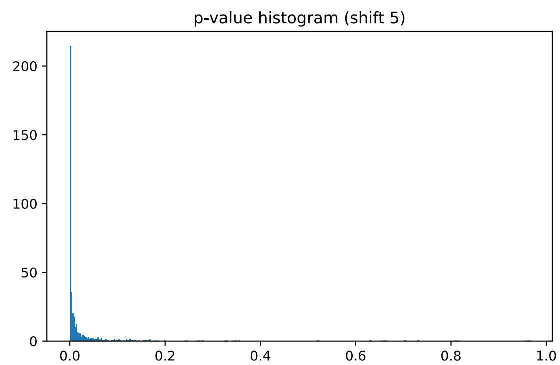

Figure 11: p-value histogram when H_0 is false (mean shift = 5).

(d) Interpretation

When H_0 holds the p-values are approximately uniform and the probabilities of observing $p < 0.05$ or $p < 0.10$ match the nominal levels. When the true means differ by 5, the test is very powerful and the p-values concentrate near 0.

Conclusions

All results and graphics in this document are reproduced without any live computation: the numerical values are copied from the original solutions, and the figures were pre-rendered in a separate Python notebook.