1 Expectation

Suppose $X$ is a uniformly distributed \textit{integer} between 1 and 6. You can think of $X$ as the result of throwing a six-sided die.

a. What is the probability distribution of $X$? Generate 100 samples from $X$, make a histogram of their relative frequency and verify that they match your probability distributions. Try for 1000 samples and 100,000 samples.

b. Calculate the expected value of $X$. Draw 100 samples from $X$, and calculate the empirical mean. Is it equal to the expected value? Why/why not?

c. Draw 10,000 samples from $X$, and calculate the average of the first 100, the average of the first 200, the average of the first 300, etc. Plot the average values against the number of samples, and plot a horizontal dashed line indicating the expected value. Describe in words how the expected value and the sample averages relate to each other.

Suppose now that $Y$ is a uniformly distributed integer between 0 and 4, and $Z = X + Y$.

d. Calculate the probability distribution of $Z$, by convolving the probability distributions of $X$ and $Y$. Write some code that converts samples from $X$ and $Y$ into samples from $Z$ (Hint: you only need 1 character). Verify that a histogram of your samples matches the probability distribution you’ve calculated.

e. Calculate the expected value of $Z$, and the empirical mean of a bunch of samples. Show that, if you gather enough samples, the empirical mean converges to the expected value. How does the expected value of $Z$ relate to that of $X$ and $Y$?

2 Confidence intervals

a. Write some code to generate coin tosses from an unfair coin. Generate 30 samples with a coin which has a 55% chance of landing on heads.

b. Pretend you forgot whether your coin was fair or not and you want to find this out from the data you just generated. Calculate the likelihood of the data as a function of the hypothesized coin fairness.

c. Give a 90% confidence interval for the fairness of this coin. Does a fair coin lie within the confidence interval?

d. Calculate the P-value of your data: generate batches of 20 samples from a fair coin (the null hypothesis), and compute your test statistic (the number of heads) for each
of these batches. Build the cumulative probability distribution of the test statistic on these batches, and find out where in that distribution your real data lies.

e. Which approach to inferring fairness do you prefer? c or d?

f. Optional: re-run a-d with 10, 20, 50 and 100 samples. How many samples do you need to reliably infer the coin to be unfair?
3 Conditional probability

a. Look around you, and count how many people in the room are female/male, how many wear glasses, and all 4 combinations of these two. Write some code to compute the conditional probabilities: \( p(\text{female}|\text{glasses}) \), \( p(\text{male}|\text{no glasses}) \), \( p(\text{glasses}|\text{female}) \), \( p(\text{no glasses}|\text{male}) \). Every time someone walks in or out of the room, modify your code accordingly.

b. Suppose you have a diagnostic test for a nasty but rare disease: linearalgebritis (note that the only cure is to take a remedial linear algebra class every year for the rest of your life). Only 0.1% of the population have this disease. Your test has a false positive rate of 1% (meaning that the probability of a positive result in a healthy patient is 1%) and a hit rate of 95% (the probability of a positive result in a patient suffering from linearalgebritis is 95%). Use Bayes’ rule to compute the probability for a patient to have linearalgebritis given that the diagnostic test comes back positive. What if you re-run the test, and it’s still positive?

4 Method of moments

The \( n \)-th moment of a probability distribution \( p(x) \) is the expected value of \( x^n \).

a. Calculate the third moment of of binomial distribution: \( p(k) = \binom{N}{k} p^k (1-p)^{N-k} \) with \( N = 6 \) and \( p = 0.4 \). Calculate the fourth moment of \( Z \) in the previous exercise.

b. What are the first and second moment of a Gaussian with mean \( \mu \) and standard deviation \( \sigma \)? (don’t do the integrals, use your intuition - or Google). Verify your expression by drawing samples and calculating the empirical mean.

The idea of the method of moments is that you take a probability distribution \( p(x|\theta) \) with some parameters \( \theta \). You calculate the first, second, third, etc. moments as a function of \( \theta \). You then invert this function, expressing the parameters \( \theta \) as a function of the moments. Finally, you replace the moments in your expression by the empirical moments, and that’s your estimator.

d. What is the method-of-moments estimator for the mean and standard deviation of a Gaussian? How many moments do you need to calculate?

c. Let’s say I told you to calculate the method-of-moments estimator of the ‘Fréchet distribution’: \( p(x) = \frac{a}{s} \left( \frac{x-m}{s} \right)^{-1-a} \exp \left( - \left( \frac{x-m}{s} \right)^{-a} \right) \). How many moments of this distribution should I calculate?
5 The Laplace distribution

In this question, you’re going to perform inference with the probability distribution

\[ p(x) = \frac{1}{2\lambda} \exp\left(-\frac{|x - \mu|}{\lambda}\right) \]

This distribution has mean \( \mu \) and standard deviation \( \lambda \sqrt{2} \).

a. Choose values for \( \mu \) and \( \lambda \), and plot the probability density function of the Laplace distribution. In the same figure, plot the probability density function of a Gaussian with the same mean and standard deviation. Describe how they’re different.

b. Write some code to generate samples from the Laplace distribution. Hint: use Matlab’s built-in function \texttt{exprnd}. Verify that the mean and standard deviation of these samples are indeed equal to \( \mu \) and \( \lambda \sqrt{2} \), respectively.

c. Let’s do inference: draw 30 samples \( x_1, \ldots, x_{30} \) from the Laplace distribution. Pretend that you forgot what \( \mu \) was (but you remember \( \lambda \)) and you want to estimate it from your data. Compute the likelihood of your data given a hypothesized value for \( \mu \). Plot this likelihood for a range of hypothesized \( \mu \). In a separate figure, plot the logarithm of this likelihood function. Write some code to compute the maximum-likelihood estimate of \( \mu \) given \( x_1, \ldots, x_{30} \).

d. Generate batches of 30 samples each from the Laplace distribution, and show that for each batch, the maximum-likelihood estimate of \( \mu \) is equal to the median of \( x_1, \ldots, x_{30} \).

6 Gaussian estimation

a. Write a function which takes as input a vector of data \( x_1, \ldots, x_N \), a prior over \( \mu \) and \( \sigma \), and returns the posterior distribution as well as maximum-likelihood estimates \( \hat{\mu} \) and \( \hat{\sigma} \).

b. Verify that your function is correct by giving it a flat prior, and show that the maximum-likelihood estimates matches the formula’s derived in class.

c. Write a function which takes as input a posterior distribution over \( \mu \) and \( \sigma \), one more data point \( x \), and returns the new posterior distribution as well as maximum-likelihood estimates \( \hat{\mu} \) and \( \hat{\sigma} \).
7 Kalman filter

For this problem, we’re studying time series $\vec{x}$ and $\vec{z}$ generated by a probabilistic process:

$$
\begin{align*}
    x_t & \sim \mathcal{N}(x_{t-1}, \sigma_x^2) \\
    z_t & \sim \mathcal{N}(x_t, \sigma_z^2)
\end{align*}
$$

a. Generate $\vec{x}$ and $\vec{z}$ given $\sigma_x$ and $\sigma_z$. Fix $x_0 = 0$. If you can, vectorize your code. Hint: use \texttt{cumsum}.

b. We now imagine that you observe $\vec{z}$, but not $\vec{x}$. It’s a hidden, or latent state. You do know $\sigma_x$ and $\sigma_z$. Write a function to compute the posterior distribution over the latent state $x_t$ given $z_t$ and $x_{t-1}$.

c. Write a function to compute the posterior distribution over the latent state $x_t$ given $z_t$ and a distribution over $x_{t-1}$.

d. Use the above functions to compute the posterior over $x_t$ at all time points, and plot the data $\vec{z}$ as well as the maximum-likelihood estimate $\hat{\vec{x}}$. How does the estimate for $\vec{x}$ relate to the data $\vec{z}$?