1 Predictable Gaussians

Note 21

Let Y be the result of a fair coin flip, and X be a normally distributed random variable with parameters dependent on Y. That is, if $Y = 1$, then $X \sim N(\mu_1, \sigma_1^2)$, and if $Y = 0$, then $X \sim N(\mu_0, \sigma_0^2)$.

(a) Sketch the two distributions of X overlaid on the same graph for the following cases:

(i) $\sigma_0^2 = \sigma_1^2, \mu_0 \neq \mu_1$

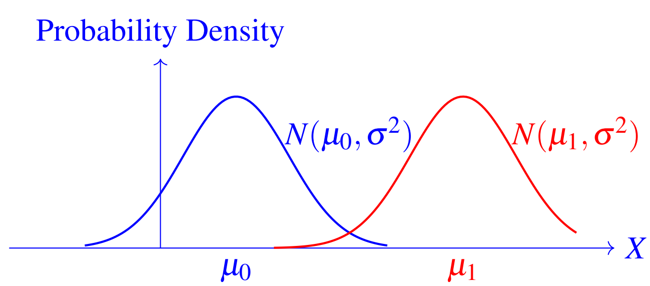
(ii) $\sigma_0^2 \neq \sigma_1^2, \mu_0 = \mu_1$

(b) Bayes' rule for mixed distributions can be formulated as $\mathbb{P}[Y = 1 \mid X = x] = \frac{\mathbb{P}[Y=1] f_{X|Y=1}(x)}{f_X(x)}$ where Y is a discrete distribution and X is a continuous distribution. Compute $\mathbb{P}[Y = 1 \mid X = x]$, and show that this can be expressed in the form of $\frac{1}{1+e^{\gamma}}$ for some expression γ. (Hint: any value z can be equivalently expressed as $e^{\ln(z)}$)

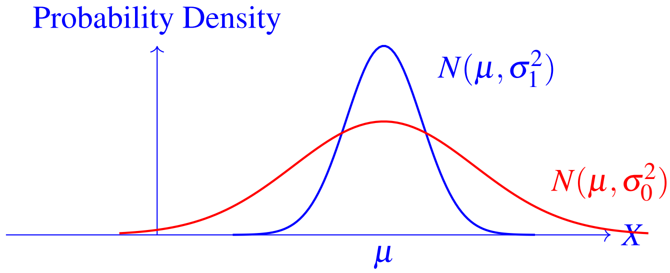
(c) In the special case where $\sigma_0^2 = \sigma_1^2$ find a simple expression for the value of x where $\mathbb{P}[Y = 1 \mid X = x] = \mathbb{P}[Y = 0 \mid X = x] = 1/2$, and interpret what the expression represents. (Hint: the identity $(a+b)(a-b) = a^2 - b^2$ may be useful)

Solution:

(a) (i) In this case, there are two bell curves with the same spread/width due to the variances being equal, but being centered at different means.



(ii) In this case, there will be two bell curves centered at the same mean, but the one with lower variance will be skinnier and taller, due to more of the probability density being centered closer to the mean.



(b)

$$\begin{aligned}
&\mathbb{P}[Y = 1 \mid X = x] \\
&= \frac{\mathbb{P}[Y = 1] f_{X|Y=1}(x)}{\mathbb{P}[Y = 1] f_{X|Y=1}(x) + \mathbb{P}[Y = 0] f_{X|Y=0}(x)} \\
&= \frac{\frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right)}{\frac{1}{\sqrt{2\pi\sigma_1^2}} \exp\left(-\frac{(x-\mu_1)^2}{2\sigma_1^2}\right) + \frac{1}{\sqrt{2\pi\sigma_0^2}} \exp\left(-\frac{(x-\mu_0)^2}{2\sigma_0^2}\right)} \\
&= \frac{1}{1 + \frac{\sigma_1}{\sigma_0} \exp\left(\frac{(x-\mu_1)^2}{2\sigma_1^2} - \frac{(x-\mu_0)^2}{2\sigma_0^2}\right)} \\
&= \frac{1}{1 + \exp\left(\ln\left(\frac{\sigma_1}{\sigma_0}\right) + \frac{(x-\mu_1)^2}{2\sigma_1^2} - \frac{(x-\mu_0)^2}{2\sigma_0^2}\right)}.
\end{aligned}$$

Which is of the desired form, with $\gamma = \ln\left(\frac{\sigma_1}{\sigma_0}\right) + \left(\frac{(x-\mu_1)^2}{2\sigma_1^2} - \frac{(x-\mu_0)^2}{2\sigma_0^2}\right)$

(c) Note that $\mathbb{P}[Y = 1 \mid X = x] = \frac{1}{2}$ implies that $\exp(\gamma) = 1$, which means that $\gamma = 0$. Thus, $\ln\left(\frac{\sigma_1}{\sigma_0}\right) + \left(\frac{(x-\mu_1)^2}{2\sigma_1^2} - \frac{(x-\mu_0)^2}{2\sigma_0^2}\right) = 0$. Using the conditions from the problem statement, we can simplify this expression.

$$\begin{aligned}
&\ln\left(\frac{\sigma_1}{\sigma_0}\right) + \left(\frac{(x-\mu_1)^2}{2\sigma^2} - \frac{(x-\mu_0)^2}{2\sigma^2}\right) = 0 \\
&0 + \left(\frac{(x-\mu_1)^2}{2\sigma^2} - \frac{(x-\mu_0)^2}{2\sigma^2}\right) = 0 \\
&(x-\mu_1)^2 = (x-\mu_2)^2 \\
&x^2 - 2\mu_1 x + \mu_1^2 = x^2 - 2\mu_2 x + \mu_2^2 \\
&2(\mu_2 - \mu_1)x = \mu_2^2 - \mu_1^2 \\
&x = \frac{\mu_2^2 - \mu_1^2}{2(\mu_2 - \mu_1)} = \frac{\mu_2 + \mu_1}{2}
\end{aligned}$$

Notice that x becomes the average, or center, of the two means.

2 Moments of the Gaussian

Note 21 For a random variable X, the quantity $\mathbb{E}[X^k]$ for $k \in \mathbb{N}$ is called the *kth moment* of the distribution. In this problem, we will calculate the moments of a standard normal distribution.

(a) Prove the identity

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{tx^2}{2}\right) \mathrm{d}x = t^{-1/2}$$

for $t > 0$.

Hint: Consider a normal distribution with variance $\frac{1}{t}$ and mean 0.

(b) For the rest of the problem, X is a standard normal distribution (with mean 0 and variance 1). Use part (a) to compute $\mathbb{E}[X^{2k}]$ for $k \in \mathbb{N}$.

Hint: Try differentiating both sides with respect to t, k times. You may use the fact that we can differentiate under the integral without proof.

(c) Compute $\mathbb{E}[X^{2k+1}]$ for $k \in \mathbb{N}$.

Solution:

(a) Note that a normal distribution with mean 0 and variance t^{-1} has the density function

$$f(x) = \frac{\sqrt{t}}{\sqrt{2\pi}} \exp\left(-\frac{tx^2}{2}\right),$$

and since the density must integrate to 1, we see that

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{tx^2}{2}\right) \mathrm{d}x = t^{-1/2}.$$

(b) Differentiating the identity from (a) k times with respect to t, we obtain a LHS of

$$\begin{aligned}
\frac{\mathrm{d}^k}{\mathrm{d}t^k}\left[\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{tx^2}{2}\right) \mathrm{d}x\right] &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \frac{\mathrm{d}^k}{\mathrm{d}t^k}\left[\exp\left(-\frac{tx^2}{2}\right)\right] \mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} (-1)^k \frac{x^{2k}}{2^k} \exp\left(-\frac{tx^2}{2}\right) \mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \frac{(-1)^k}{2^k} \int_{-\infty}^{\infty} x^{2k} \exp\left(-\frac{tx^2}{2}\right) \mathrm{d}x
\end{aligned}$$

Here, we use the fact that everything involving x is a constant with respect to t.

Looking at the RHS, we have

$$\frac{\mathrm{d}^k}{\mathrm{d}t^k}\left[t^{-1/2}\right] = (-1)^k \frac{1 \cdot 3 \cdots (2k-3) \cdot (2k-1)}{2^k} t^{-(2k+1)/2}$$

Together, this means that

$$\frac{1}{\sqrt{2\pi}}\frac{\cancel{(-1)^k}}{\cancel{2^k}}\int_{-\infty}^{\infty} x^{2k}\exp\left(-\frac{tx^2}{2}\right)\,\mathrm{d}x = \cancel{(-1)^k}\frac{1\cdot 3\cdots(2k-3)\cdot(2k-1)}{\cancel{2^k}}t^{-(2k+1)/2}$$

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^{2k}\exp\left(-\frac{tx^2}{2}\right)\,\mathrm{d}x = (1\cdot 3\cdots(2k-3)\cdot(2k-1))t^{-(2k+1)/2}$$

If we set $t=1$, we get

$$\mathbb{E}[X^{2k}] = \int_{-\infty}^{\infty} x^{2k}\cdot\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)\,\mathrm{d}x = \prod_{i=1}^{k}(2i-1).$$

This is sometimes denoted $(2k-1)!!$. Note that we can also write the result as

$$\mathbb{E}[X^{2k}] = (2k-1)!! = \frac{(2k)!}{2\cdot 4\cdots(2k-2)\cdot(2k)} = \frac{(2k)!}{2^k k!}.$$

(c) $\mathbb{E}[X^{2k+1}] = 0$, since the density function is symmetric around 0.

3 Chebyshev's Inequality vs. Central Limit Theorem

Note 17
Note 21

Let n be a positive integer. Let $X_1, X_2, \ldots, X_n$ be i.i.d. random variables with the following distribution:

$$\mathbb{P}[X_i = -1] = \frac{1}{12}; \qquad \mathbb{P}[X_i = 1] = \frac{9}{12}; \qquad \mathbb{P}[X_i = 2] = \frac{2}{12}.$$

(a) Calculate the expectations and variances of X_1, $\sum_{i=1}^n X_i$, $\sum_{i=1}^n (X_i - \mathbb{E}[X_i])$, and

$$Z_n = \frac{\sum_{i=1}^n (X_i - \mathbb{E}[X_i])}{\sqrt{n/2}}.$$

(b) Use Chebyshev's Inequality to find an upper bound b for $\mathbb{P}[|Z_n| \geq 2]$.

(c) Use b from the previous part to bound $\mathbb{P}[Z_n \geq 2]$ and $\mathbb{P}[Z_n \leq -2]$.

(d) As $n \to \infty$, what is the distribution of Z_n?

(e) We know that if $Z \sim \mathcal{N}(0,1)$, then $\mathbb{P}[|Z| \leq 2] = \Phi(2) - \Phi(-2) \approx 0.9545$. As $n \to \infty$, provide approximations for $\mathbb{P}[Z_n \geq 2]$ and $\mathbb{P}[Z_n \leq -2]$.

Solution:

(a) Firstly, let us calculate $\mathbb{E}[X_1]$ and $\mathrm{Var}(X_1)$; we have

$$\mathbb{E}[X_1] = -\frac{1}{12} + \frac{9}{12} + \frac{4}{12} = 1$$

$$\mathrm{Var}(X_1) = \frac{1}{12}\cdot 2^2 + \frac{9}{12}\cdot 0^2 + \frac{2}{12}\cdot 1^2 = \frac{1}{2}.$$

Using linearity of expectation and variance (since $X_1, \dots, X_n$ are independent), we find that

$$\mathbb{E}\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n \mathbb{E}[X_i] = n$$
$$\mathrm{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{Var}(X_i) = \frac{n}{2}$$

Again, by linearity of expectation,

$$\mathbb{E}\left[\sum_{i=1}^n (X_i - \mathbb{E}[X_i])\right] = \mathbb{E}\left[\sum_{i=1}^n X_i - n\right] = n - n = 0.$$

Subtracting a constant does not change the variance, so

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

as before.

Using the scaling properties of the expectation and variance, we finally have

$$\mathbb{E}[Z_n] = \mathbb{E}\left[\frac{\sum_{i=1}^n (X_i - \mathbb{E}[X_i])}{\sqrt{n/2}}\right] = \frac{1}{\sqrt{n/2}}\mathbb{E}\left[\sum_{i=1}^n (X_i - \mathbb{E}[X_i])\right] = \frac{0}{\sqrt{n/2}} = 0$$
$$\mathrm{Var}(Z_n) = \mathrm{Var}\left(\frac{\sum_{i=1}^n (X_i - \mathbb{E}[X_i])}{\sqrt{n/2}}\right) = \frac{1}{n/2}\mathrm{Var}\left(\sum_{i=1}^n (X_i - \mathbb{E}[X_i])\right) = \frac{n/2}{n/2} = 1$$

(b) Using Chebyshev's, we have

$$\mathbb{P}[|Z_n| \geq 2] \leq \frac{\mathrm{Var}(Z_n)}{2^2} = \frac{1}{4}$$

since $\mathbb{E}[Z_n] = 0$ and $\mathrm{Var}(Z_n) = 1$ as we computed in the previous part.

(c) $\frac{1}{4}$ for both, since we have

$$\mathbb{P}[Z_n \geq 2] \leq \mathbb{P}[|Z_n| \geq 2]$$
$$\mathbb{P}[Z_n \leq -2] \leq \mathbb{P}[|Z_n| \geq 2]$$

(d) By the Central Limit Theorem, we know that $Z_n \to \mathcal{N}(0,1)$, the standard normal distribution.

(e) Since $Z_n \to \mathcal{N}(0,1)$, we can approximate $\mathbb{P}[|Z_n| \geq 2] \approx 1 - 0.9545 = 0.0455$. By the symmetry of the normal distribution, $\mathbb{P}[Z_n \geq 2] = \mathbb{P}[Z_n \leq -2] \approx 0.0455/2 = 0.02275$.

It is interesting to note that the CLT provides a much smaller answer than Chebyshev. This is due to the fact that the CLT is applied to a particular kind of random variable, namely the (scaled) sum of a bunch of random variables. Chebyshev's inequality, however, holds for any random variable, and is therefore weaker.

4 (Optional) LLSE and Graphs

Note 20

Consider a graph with n vertices numbered 1 through n, where n is a positive integer ≥ 2. For each pair of distinct vertices, we add an undirected edge between them independently with probability p. Let D_1 be the random variable representing the degree of vertex 1, and let D_2 be the random variable representing the degree of vertex 2.

(a) Compute $\mathbb{E}[D_1]$ and $\mathbb{E}[D_2]$.

(b) Compute $\mathrm{Var}(D_1)$.

(c) Compute $\mathrm{cov}(D_1, D_2)$.

(d) Using the information from the first three parts, what is $L(D_2 \mid D_1)$?

Solution:

Throughout this problem, let $X_{i,j}$ be an indicator random variable for whether the edge between vertex i and vertex j exists, for $i, j = 1, \ldots, n$. Note that $X_{i,j} = X_{j,i}$.

(a) Observing that $D_1, D_2 \sim \text{Binomial}(n-1, p)$, we obtain $\mathbb{E}[D_1] = \mathbb{E}[D_2] = (n-1)p$.

Anyway, it is good to review how we derived the expectation of the binomial distribution in the first place. By linearity of expectation,

$$\mathbb{E}[D_1] = \mathbb{E}\Big[\sum_{i=2}^{n} X_{1,j}\Big] = \sum_{i=2}^{n} \mathbb{E}[X_{1,j}] = (n-1)\,\mathbb{E}[X_{i,j}] = (n-1)p.$$

By symmetry, $\mathbb{E}[D_2] = (n-1)p$ also.

(b) Since $D_1, D_2 \sim \text{Binomial}(n-1, p)$, then $\mathrm{Var}\,D_1 = \mathrm{Var}\,D_2 = (n-1)p(1-p)$.

Again, it is good to review how we calculated the variance of the binomial distribution.

Solution 1: Write the variance of D_1 as a sum of covariances.

$$\begin{aligned}
\mathrm{Var}(D_1) &= \mathrm{cov}\Big(\sum_{i=2}^{n} X_{1,i}, \sum_{i=2}^{n} X_{1,i}\Big) = (n-1)\,\mathrm{Var}(X_{1,2}) + \big((n-1)^2 - (n-1)\big)\,\mathrm{cov}(X_{1,2}, X_{1,3}) \\
&= (n-1)p(1-p) + 0 = (n-1)p(1-p).
\end{aligned}$$

We used the fact that $X_{1,i}$ and $X_{1,j}$ are independent if $i \neq j$, so their covariance is zero.

Solution 2: Compute the variance directly.

$$\begin{aligned}
\mathrm{Var}(X) &= \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \mathbb{E}\Big[\Big(\sum_{i=2}^{n} X_{1,i}\Big)^2\Big] - (n-1)^2 p^2 \\
&= (n-1)\,\mathbb{E}[X_{1,2}^2] + \big((n-1)^2 - (n-1)\big)\,\mathbb{E}[X_{1,2}X_{1,3}] - (n-1)^2 p^2 \\
&= (n-1)p + (n^2 - 3n + 2)p^2 - (n-1)^2 p^2 \\
&= (n-1)p + (n-1)(n-2)p^2 - (n-1)^2 p^2 = (n-1)p\big(1 + (n-2)p - (n-1)p\big) \\
&= (n-1)p(1-p)
\end{aligned}$$

(c) We can write

$$\mathrm{cov}(D_1, D_2) = \mathrm{cov}\Big(\sum_{i=2}^{n} X_{1,i}, \sum_{i=1, i\neq 2}^{n} X_{2,i}\Big) = \sum_{i=2}^{n} \sum_{j=1, j\neq 2}^{n} \mathrm{cov}(X_{1,i}, X_{2,j}).$$

Note that all pairs of $X_{1,i}, X_{2,j}$ are independent except for when $i = 2$ and $j = 1$, so all terms in the sum are zero except for $\mathrm{cov}(X_{1,2}, X_{2,1})$, and our covariance is just equal to $\mathrm{cov}(X_{1,2}, X_{2,1}) = \mathrm{Var}(X_{1,2}) = p(1-p)$.

(d) Since

$$L(D_2 \mid D_1) = \mathbb{E}[D_2] + \frac{\mathrm{cov}(D_1, D_2)}{\mathrm{Var}(D_1)}(D_1 - \mathbb{E}[D_1]),$$

we plug in our values from the first three parts to get that

$$\begin{aligned}
L(D_2 \mid D_1) &= (n-1)p + \frac{p(1-p)}{(n-1)p(1-p)}\big(D_1 - (n-1)p\big) \\
&= (n-1)p + \frac{1}{n-1}\big(D_1 - (n-1)p\big) = \frac{1}{n-1}D_1 + (n-2)p.
\end{aligned}$$