1 Interesting Gaussians

(a) If $X \sim N(0, \sigma_X^2)$ and $Y \sim N(0, \sigma_Y^2)$ are independent, then what is $\mathbb{E}\left[(X+Y)^k\right]$ for any *odd* $k \in \mathbb{N}$?

(b) Let $f_{\mu,\sigma}(x)$ be the density of a $N(\mu, \sigma^2)$ random variable, and let X be distributed according to $\alpha f_{\mu_1,\sigma_1}(x) + (1-\alpha) f_{\mu_2,\sigma_2}(x)$ for some $\alpha \in [0,1]$. Compute $\mathbb{E}[X]$ and $\operatorname{Var}(X)$. Is X normally distributed?

Solution:

(a) $\mathbb{E}\left[(X+Y)^k\right] = 0$.

Since X and Y are Gaussians, so must $Z = X+Y$ be. Moreover, as Z is of mean 0, we know that its distribution f_Z is symmetric around the origin, i.e. $f_Z(x) = f_Z(-x)$ for any $a, b \in \mathbb{R}$. Therefore,

$$\begin{aligned}
\mathbb{E}\left[(X+Y)^k\right] = \mathbb{E}\left[Z^k\right] &= \int_{-\infty}^{\infty} x^k f_Z(x)\,\mathrm{d}x \\
&= \int_{-\infty}^{0} x^k f_Z(x)\,\mathrm{d}x + \int_{0}^{\infty} x^k f_Z(x)\,\mathrm{d}x \\
&= \int_{0}^{\infty} (-x)^k f_Z(-x)\,\mathrm{d}x + \int_{0}^{\infty} x^k f_Z(x)\,\mathrm{d}x \\
&= -\int_{0}^{\infty} x^k f_Z(x)\,\mathrm{d}x + \int_{0}^{\infty} x^k f_Z(x)\,\mathrm{d}x \\
&= 0,
\end{aligned}$$

since k is odd.

(b) $\mathbb{E}[X] = \alpha\mu_1 + (1-\alpha)\mu_2$, $\operatorname{Var}(X) = \alpha\left(\sigma_1^2 + \mu_1^2\right) + (1-\alpha)\left(\sigma_2^2 + \mu_2^2\right) - (\mathbb{E}[X])^2$. No, X is not necessarily normally distributed.

$$\begin{aligned}
\mathbb{E}[[]X] := \mu &= \int_{-\infty}^{\infty} x\left(\alpha f_{\mu_1,\sigma_1}(x) + (1-\alpha) f_{\mu_2,\sigma_2}(x)\right)\mathrm{d}x \\
&= \alpha \int_{-\infty}^{\infty} x f_{\mu_1,\sigma_1}(x)\,\mathrm{d}x + (1-\alpha)\int_{-\infty}^{\infty} x f_{\mu_2,\sigma_2}(x)\,\mathrm{d}x = \alpha\mu_1 + (1-\alpha)\mu_2 \\
\operatorname{Var}(X) := \sigma^2 = \mathbb{E}\left[X^2\right] - \mu^2 &= \alpha \int_{-\infty}^{\infty} x^2 f_{\mu_1,\sigma_1}(x)\,\mathrm{d}x + (1-\alpha)\int_{-\infty}^{\infty} x^2 f_{\mu_2,\sigma_2}(x)\,\mathrm{d}x - \mu^2 \\
&= \alpha\left(\sigma_1^2 + \mu_1^2\right) + (1-\alpha)\left(\sigma_2^2 + \mu_2^2\right) - \mu^2.
\end{aligned}$$

We know that the density of $N(\mu,\sigma)$ has a unique maximum at $x=\mu$; however, if, e.g. $\alpha = 1/2, \mu_1 = -10, \mu_2 = 10, \sigma_1 = \sigma_2 = 1$, then $\alpha f_{\mu_1,\sigma_1} + (1-\alpha) f_{\mu_2,\sigma_2}$ has two maxima, and so cannot be the density of a Gaussian.

2 Continuous Joint Densities

The joint probability density function of two random variables X and Y is given by $f(x,y) = Cxy$ for $0 \leq x \leq 1, 0 \leq y \leq 2$, and 0 otherwise (for a constant C).

(a) Find the constant C that ensures that $f(x,y)$ is indeed a probability density function.

(b) Find $f_X(x)$, the marginal distribution of X.

(c) Find the conditional distribution of Y given $X = x$.

(d) Are X and Y independent?

Solution:

(a) Since $f(x,y)$ is a probability density function, it must integrate to 1. Then:

$$1 = \int_0^1 \int_0^2 Cxy\,dy\,dx = \int_0^1 2Cx dx = C$$

Therefore, $C = 1$.

(b) To get the marginal distribution of X, we integrate the joint distribution with respect to Y. So:

$$f_X(x) = \int_0^2 f(x,y)dy = \int_0^2 xydy = 2x$$

This is the marginal distribution for $0 \leq x \leq 1$.

(c) The conditional distribution of Y given by

$$f_{Y|X}(y|x) = \frac{f(x,y)}{f_X(x)} = \frac{xy}{2x} = \frac{y}{2}$$

(d) The conditional distribution of Y given $X = x$ does not depend on x, so they are independent.

Alternatively, you could find the marginal distribution of Y and see it is the same as the conditional distribution of Y:

$$f_Y(y) = \int_0^1 f(x,y)dx = \int_0^1 xydx = \frac{y}{2}$$

Notice that since X and Y are independent, $f_X(x)f_Y(y) = xy = f_{X,Y}(x,y)$, i.e. the product of the marginal distributions is the same as the joint distribution.

3 Binomial Concentration

Here, we will prove that the binomial distribution is *concentrated* about its mean as the number of trials tends to ∞. Suppose we have i.i.d. trials, each with a probability of success $1/2$. Let S_n be the number of successes in the first n trials (n is a positive integer), and define

$$Z_n := \frac{S_n - n/2}{\sqrt{n}/2}.$$

(a) What are the mean and variance of Z_n?

(b) What is the distribution of Z_n as $n \to \infty$?

(c) Use the bound $\mathbb{P}[Z > z] \le (\sqrt{2\pi}z)^{-1}\,\mathrm{e}^{-z^2/2}$ when Z is a standard normal in order to approximately bound $\mathbb{P}[S_n/n > 1/2 + \delta]$, where $\delta > 0$.

Solution:

(a) 0 and 1, respectively. We made them so, in order to apply the CLT. Here are the computations.

$$\mathbb{E}[Z_n] = \frac{1}{\sqrt{n}/2}\mathbb{E}\left[S_n - \frac{n}{2}\right] = \frac{1}{\sqrt{n}/2}\left(\mathbb{E}[S_n] - \frac{n}{2}\right) = 0,$$
$$\operatorname{Var}(Z_n) = \frac{1}{n/4}\operatorname{Var}\left(S_n - \frac{n}{2}\right) = \frac{1}{n/4}\operatorname{Var} S_n = 1,$$

since $S_n \sim \text{Binomial}(n, 1/2)$.

(b) The CLT tells us that $Z_n \to \mathcal{N}(0,1)$.

(c) In order to apply the bound, we must apply it to Z_n.

$$\mathbb{P}\left[\frac{S_n}{n} > \frac{1}{2} + \delta\right] = \mathbb{P}\left[\frac{S_n - n/2}{n} > \delta\right] = \mathbb{P}\left[\frac{S_n - n/2}{\sqrt{n}/2} > 2\delta\sqrt{n}\right] \approx \mathbb{P}[Z_n > 2\delta\sqrt{n}]$$
$$\le \frac{1}{2^{3/2}\delta\sqrt{\pi n}} e^{-2\delta^2 n}$$