DIS 12B

1 Sum of Independent Gaussians

In this question, we will introduce an important property of the Gaussian distribution: the sum of independent Gaussians is also a Gaussian.

Let X and Y be independent standard Gaussian random variables. Recall that the density of the standard Gaussian is

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

(a) What is the joint density of X and Y?

(b) Observe that the joint density of X and Y, $f_{X,Y}(x,y)$, only depends on the quantity x^2+y^2, which is the distance from the origin. In other words, the Gaussian is *rotationally symmetric*. Next, we will try to find the density of $X+Y$. To do this, draw a picture of the Cartesian plane and draw the region $x+y \le c$, where c is a real number of your choice.

(c) Now, rotate your picture clockwise by $\pi/4$ so that the line $X+Y=c$ is now vertical. Redraw your figure. Let X' and Y' denote the random variables which correspond to the $\pi/4$ clockwise rotation of (X,Y) and express the new shaded region in terms of X' and Y'.

(d) By rotational symmetry of the Gaussian, (X',Y') has the same distribution as (X,Y). Argue that $X+Y$ has the same distribution as $\sqrt{2}Z$, where Z is a standard Gaussian. This proves the following important fact: *the sum of independent Gaussians is also a Gaussian*. Notice that $X \sim \mathcal{N}(0,1)$, $Y \sim \mathcal{N}(0,1)$ and $Z \sim \mathcal{N}(0,2)$. *In general, if X and Y are independent Gaussians, then $X+Y$ is a Gaussian with mean $\mu_X+\mu_Y$ and variance $\sigma_X^2+\sigma_Y^2$.*

(e) Recall the CLT:

> If $\{X_i\}_{i\in\mathbb{N}}$ is a sequence of i.i.d. random variables with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 < \infty$, then:
>
> $$\frac{X_1+\cdots+X_n-n\mu}{\sigma\sqrt{n}} \xrightarrow{\text{in distribution}} \mathcal{N}(0,1) \qquad \text{as } n \to \infty.$$

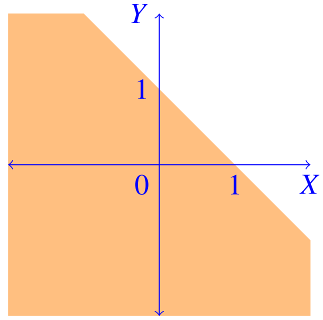
Prove that the CLT holds for the special case when the X_i are i.i.d. $\mathcal{N}(0,1)$.

Solution:

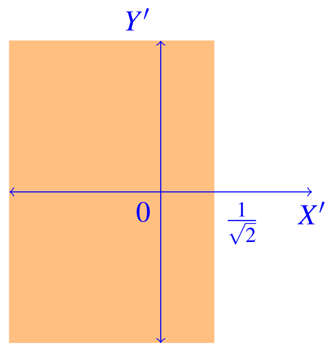
(a) By independence, we have

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) = \frac{1}{2\pi}\exp\Big(-\frac{x^2+y^2}{2}\Big).$$

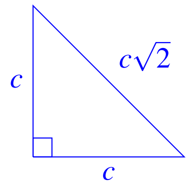
(b) We draw the line for $c = 1$.



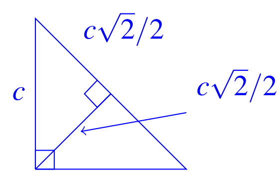
(c) Here is the new figure after the rotation (for $c = 1$).



For general $c \in \mathbb{R}$, the new region is $\{X' \le c/\sqrt{2}\}$. To see why, draw the triangle: We want



to find the distance between the origin and the long side of the triangle, and we can do so by adding a diagonal:



(d) We observe that $\mathbb{P}(X+Y \le c) = \mathbb{P}(X' \le c/\sqrt{2}) = \mathbb{P}(\sqrt{2}X' \le c)$, where X' is a standard Gaussian, so this proves the claim.

(e) Here, $\mu = 0$ and $\sigma = 1$. So, by the previous part,

$$\frac{X_1 + \cdots + X_n}{\sqrt{n}} \sim \frac{1}{\sqrt{n}} \mathcal{N}(0,n) \sim \mathcal{N}(0,1).$$

2 Inequality Practice

(a) X is a random variable such that $X > -5$ and $\mathbb{E}[X] = -3$. Find an upper bound for the probability of X being greater than or equal to -1.

(b) You roll a die 100 times. Let Y be the sum of the numbers that appear on the die throughout the 100 rolls. Use Chebyshev's inequality to bound the probability of the sum Y being greater than 400 or less than 300.

Solution:

(a) We want to use Markov's Inequality, but we remember that Markov's only works with non-negative random variables. Then, we define a new random variable $Y = X + 5$, where Y is always non-negative, so we can use Markov's on Y. By linearity of expectation, $\mathbb{E}[Y] = -3 + 5 = 2$. So, $\mathbb{P}[Y \geq 4] \leq 2/4 = 1/2$.

(b) Let Y_i be the number on the die for the ith roll, for $i = 1, \ldots, 100$. Then, $Y = \sum_{i=1}^{100} Y_i$. By linearity of expectation, $\mathbb{E}[Y] = \sum_{i=1}^{100} \mathbb{E}[Y_i]$.

$$\mathbb{E}[Y_i] = \sum_{j=1}^{6} j \cdot \mathbb{P}[Y_i = j] = \sum_{j=1}^{6} j \cdot \frac{1}{6} = \frac{1}{6} \cdot \sum_{j=1}^{6} j = \frac{1}{6} \cdot 21 = \frac{7}{2}$$

Then, we have $\mathbb{E}[Y] = 100 \cdot (7/2) = 350$.

$$\mathbb{E}[Y_i^2] = \sum_{j=1}^{6} j^2 \cdot \mathbb{P}[Y_i = j] = \sum_{j=1}^{6} j^2 \cdot \frac{1}{6} = \frac{1}{6} \cdot \sum_{j=1}^{6} j^2 = \frac{1}{6} \cdot 91 = \frac{91}{6}$$

Then, we have

$$\text{var}(Y_i) = \mathbb{E}[Y_i^2] - \mathbb{E}[Y_i]^2 = \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12},$$

so $\text{var}(Y) = 100(35/12)$.

Putting it all together, we use Chebyshev's to get

$$\mathbb{P}[|X - 350| \geq 50] \leq \frac{100(35/12)}{50^2} = \frac{7}{60}.$$

3 Poisson Confidence Interval

You collect n samples (n is a positive integer) $X_1, \ldots, X_n$, which are i.i.d. and known to be drawn from a Poisson distribution (with unknown mean). However, you have a bound on the mean: from a confidential source, you know that $\lambda \le 2$. Find a $1-\delta$ confidence interval ($\delta \in (0,1)$) for λ using Chebyshev's Inequality.

Solution:

Our estimator for λ is the sample mean $n^{-1}\sum_{i=1}^{n} X_i$. We apply Chebyshev's Inequality for $\varepsilon > 0$:

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \lambda\right| > \varepsilon\right) \le \frac{\operatorname{var}(n^{-1}\sum_{i=1}^{n} X_i)}{\varepsilon^2} = \frac{\operatorname{var}(\sum_{i=1}^{n} X_i)}{n^2\varepsilon^2} = \frac{\sum_{i=1}^{n}\operatorname{var} X_i}{n^2\varepsilon^2} = \frac{\operatorname{var} X_1}{n\varepsilon^2} = \frac{\lambda}{n\varepsilon^2}$$
$$\le \frac{2}{n\varepsilon^2}.$$

We want the probability of error to be at most δ, so we set

$$\frac{2}{n\varepsilon^2} \le \delta \implies \varepsilon \ge \sqrt{\frac{2}{n\delta}}.$$

Our $1-\delta$ confidence interval for λ is $(n^{-1}\sum_{i=1}^{n} X_i - \sqrt{2/(n\delta)}, n^{-1}\sum_{i=1}^{n} X_i + \sqrt{2/(n\delta)})$.