Exercise set 2

Exercise 1

Let $\xi : \Omega \to \mathbb{R}$ and $\eta : \Omega \to \mathbb{R}$ two random variables with joint distribution

$$P(\xi \in X, \eta \in Y) = \int_{X \times Y} dx dy \, p_{\xi\,,\eta}(x\,,y)$$

for X, Y any Borel set in $\mathbb{R}^2$ and $p_{\xi\,,\eta}$ a smooth probability density.

- Find the *marginal* probability densities $p_\xi(x)$ and $p_\eta(y)$
- Write the expression for the probability $P\,(x \,\leq\, \xi \leq\, x + dx|\eta \,\in\, Y)$.

Exercise 2

Determine which of the following functions can be thought as a probability density and explain why:

- $\sin x$ for $x \,\in\, [0, \pi]$.
- $-\ln x$ for $x \,\in\, [0, 1]$.
- $(x/\bar{x})\, e^{-x/\bar{x}}$ for for $x \,\in\, \mathbb{R}_+$.
- $(2/\sqrt{3})\, \cos x$ for $x \,\in\, [0, 2\,\pi/3]$.

Exercise 3

Consider two independent random variables ξ_1, ξ_2 each admitting a probability density on the entire real axis. What is the meaning of the average

$$I(x) = \prec \delta\,(x - \xi_1 - \xi_2) \succ$$

where δ is the Dirac-delta function? Calculate explicitly the result in the case when ξ_1, ξ_2 are Gaussian random variables.

Exercise 4

Let $\{\xi_i\}_{i=1}^n$ a finite sequence of i.i.d. Gaussian random variables with zero mean and unit variance. Compute the probability distribution of the random variable

$$\zeta = \sum_{i=1}^{n} \xi_i^2$$

Exercise 5

Prove (at least at informal level) the following theorem:

Theorem 0.1 (*Pearson*)**.** *Let* $\{\xi_i\}_{i=1}^n$ *i.i.d. random variables such that* $\xi_i \stackrel{d}{=} \xi\ \forall i$ *and*

$$\xi : \Omega \to \{a_1, \dots a_s\} \qquad \text{with} \qquad P(\xi = a_i) = p_i \quad i = 1, \dots, s$$

with $s \ll n$*. Let* $\{\nu_i\}_{i=1}^s$ *describe the empirical distribution of the sequence i.e*

$$\nu_i := \frac{1}{n} \sum_{j=1}^n H_{a_i}(\xi_j)$$

where

$$H_a(x) = \begin{cases} 1 & \text{if } x = a \\ 0 & \text{if } x \neq a \end{cases}$$

Then the random variable

$$\chi^2_{n,s-1} := \sum_{i=1}^s \frac{(\nu_i - n\,p_i)^2}{n\,p_i}$$

has in distribution a well defined limit

$$\lim_{n \uparrow \infty} \chi^2_{n,s-1} \stackrel{d}{=} \chi^2_{s-1}$$

with $\chi^2_{s-1} : \Omega \to \mathbb{R}_+$ *such that*

$$p_{\chi^2_{s-1}}(x) = \frac{x^{\frac{s-1}{2}-1} e^{-x/2}}{2^{\frac{s-1}{2}} \Gamma\left(\frac{s-1}{2}\right)}$$

Hints:

- Step 0: make sure you have done and meditated exercise 5 of the first exercise set.
- Step 1: use the central limit theorem to argue that for any $i = 1, \dots, s$

  $$\frac{(\nu_i - n\,p_i)}{\sqrt{n\,p_i}} \stackrel{n\uparrow\infty}{\to} \zeta_i \qquad \text{in distribution}$$

  where ζ_i is a Gaussian random variable and compute its mean value and variance.
- Step 2: compute the covariance matrix C_{ij} of the ζ_i's:

  $$C_{ij} := \prec \zeta_i\, \zeta_j \succ$$

  and explain why the central limit theorem cannot be applied to the sum $\sum_{i=1}^s \zeta_i^2$.
- Step 3: consider now the sequence $\{\eta_i\}_{i=1}^s$ of i.i.d. Gaussian variables and the array $(\sqrt{p_1}, \dots, \sqrt{p_s})$. Compute the covariance matrix $\tilde{C}_{ij}$

  $$\tilde{C}_{ij} = \prec \tilde{\eta}_i\, \tilde{\eta}_j \succ \qquad \& \qquad \tilde{\eta}_i = \eta_i - \sqrt{p_i} \sum_{j=1}^s \eta_j \sqrt{p_j}$$

- Step 4: think of

$$\begin{bmatrix} \eta_1 \\ \vdots \\ \eta_s \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} \tilde{\eta}_1 \\ \vdots \\ \tilde{\eta}_s \end{bmatrix}$$

  as vectors in $\mathbb{R}^s$ and explain what is their geometrical relation. Hint: set first $s = 2$.

- Step 5: contrast C_{ij} and $\tilde{C}_{ij}$. What do you infer about the *joint* distribution of the ζ_i's?

- Step 6: use step 5 to prove

$$\sum_{i=1}^{s} \zeta_i^2 = \sum_{i=1}^{s} \tilde{\eta}_i^2$$

- Step 7: if you haven't yet, do exercise 4 above and draw your conclusions .

Remark 0.1. Pearson's theorem is the basis for the so called χ^2-test which can be applied to compare frequencies as inferred from empirical data to a guess for a discrete probability distribution which may be conjectured to provide theoretical model for the data.