Quiz

Find a match:

1. Cumulative distribution function	A. Waiting time
2. Quantile function	B. Separability
3. Pooled variance	C. Approximation
4. Exponential distribution	D. Antiderivative
5. Degrees of freedom	E. Weighted average
6. Central Limit Theorem	F. Goodness-of-fit
7. Independence	G. Constraints
8. Correlation	H. Inverse

Solution:

1. D (C.d.f. is an antiderivative of p.d.f.)
2. H (Quantile function is an inverse of c.d.f.)
3. E (Pooled variance is a weighted average of several sample variances with weighs being degrees of freedom)
4. A (Exponential distribution shows a waiting time in a Poisson process of the next event)
5. G (Degrees of freedom decrease by a number of constraints in a sample)
6. C (The CLT allows approximation of distributions in a large sample with the normal one)
7. B (Independence of random variables is defined as a separability of joint distribution into marginal ones)
8. F (Correlation is a goodness-of-fit metric for a linear regression problem)

Problem 1

A random sample of 400 married couples was selected from a large population of married couples.

- Heights of married men are approximately normally distributed with mean 70 inches and standard deviation 3 inches.
- Heights of married women are approximately normally distributed with mean 65 inches and standard deviation 2.5 inches.
- There were 20 couples in which wife was taller than her husband, and there were 380 couples in which wife was shorter than her husband.

1. Find a 95% confidence interval for the proportion of married couples in the population for which the wife is taller than her husband.
2. Suppose that a married man is selected at random and a married woman is selected at random. Find the approximate probability that the woman will be taller than the man.
3. Based on your answers to 1 and 2, are the heights of wives and their husbands independent? Explain your reasoning.

Solution:

TODO

Problem 2

Suppose 2000 points are selected independently at a random from the unit square $S = \{(x, y)\colon 0 \le x \le 1,\ 0 \le y \le 1\}$. Let W be the number of points that fall into the set $A = \{(x, y)\colon\ x^2 + y^2 < 1\}$.

1. How is W distributed?
2. Find the mean, variance and standard deviation of W.
3. Estimate probability that W is greater than 1600.

Solution:

TODO

Problem 3

Distribution of X is uniform $\mathcal{U}(-a, a)$. Sample of size $n = 2$ is available.
Consider $\widehat{a} = c \cdot (|X_1| + |X_2|)$ as a class of estimators for the parameter a.
Find c such that

1. Estimator $\widehat{a}$ is unbiased.
2. Estimator $\widehat{a}$ is the most efficient in the class. (In terms of mean square error.)

Solution:

TODO

Problem 4

Consider random variables X and Y with joint density function

$$f(x,y) = \begin{cases} \frac{1}{2} + cx, & x + y \leq 1,\ x \geq 0,\ y \geq 0, \\ 0, & \text{otherwise.} \end{cases}$$

1. Find c.
2. Find $f_X(x)$. Evaluate $\mathsf{E}(X)$.
3. Write down an expression for $f_{Y|X}(x,y)$. Find $\mathsf{E}(Y \mid X = x)$.

Solution:

TODO

Problem 5

Internal angles $\theta_1, \theta_2, \theta_3, \theta_4$ of a certain quadrilateral, located on the ground, were measured by the aerial system. It is assumed that those observations x_1, x_2, x_3, x_4 were taken with minor and independent errors, which have zero mean and identical variance σ^2.

1. Find the LSE of $\theta_1, \theta_2, \theta_3, \theta_4$.
2. Find an unbiased estimate of σ^2 in the case, described in part 1.
3. Let's assume now that the considered quadrilateral is a parallelogram with $\theta_1 = \theta_3$ and $\theta_2 = \theta_4$. How values of internal angles LSE would change? Find an unbiased estimate of σ^2 in this particular case.

Solution:

TODO

Problem 6

Suppose that student's grade for a statistics exam, X, has continuous uniform distribution at the interval $[0, 100]$. But less then 25 points means "failed", and more than 80 points is "excellent", hence the final grade Y is calculated as follows:

$$Y = \begin{cases} 0, & X < 25 \\ X, & 25 \leq X < 80 \\ 100, & X \geq 80 \end{cases}$$

1. Find c.d.f. of Y. Sketch the plot.
2. Find p.d.f. of Y. Sketch the plot.
3. Find mean and variance of X and Y.
4. Find $\mathsf{E}(Y \mid Y > 0)$.
5. Find $\mathrm{Corr}(X, Y)$.

Solution:

1. A random variable X has following p.d.f. and c.d.f.:

$$f_X(x) = \frac{1}{100} \cdot I_{\{0 \leq x \leq 100\}},$$

$$F_X(x) = \begin{cases} 1, & x \geq 100, \\ \dfrac{x}{100}, & 0 \leq x < 100, \\ 0, & x < 0. \end{cases}$$

The c.d.f. of Y coincides with that of X for $y \in [25, 80)$:

$$F_Y(y) = \frac{y}{100}, \text{ for } 25 \leq y < 80.$$

$F_Y(y)$ has a sharp increase in $y = 0$ by a value of $\mathsf{P}(X < 25) = \dfrac{1}{4}$, and is maintained on the same level up to $y = 25$:

$$F_Y(y) = \frac{1}{4}, \text{ for } 0 \leq y < 25.$$

Also, $F_Y(y)$ does not change to the right of $y = 80$ until it hits $y = 100$, where it has a sharp increase by a value of $\mathsf{P}(X \geq 80) = \frac{1}{5}$:

$$F_Y(y) = \frac{4}{5}, \text{ for } 80 \leq y < 100.$$

Thus, the c.d.f. of Y has a following view:

$$\boxed{F_Y(y) = \begin{cases} 1, & y \geq 100, \\ \frac{4}{5}, & 80 \leq y < 100, \\ \frac{y}{100}, & 25 \leq y < 80, \\ \frac{1}{4}, & 0 \leq y < 25, \\ 0, & y < 0. \end{cases}}$$

A graph of $F_Y(y)$ is shown in the fig. 1.

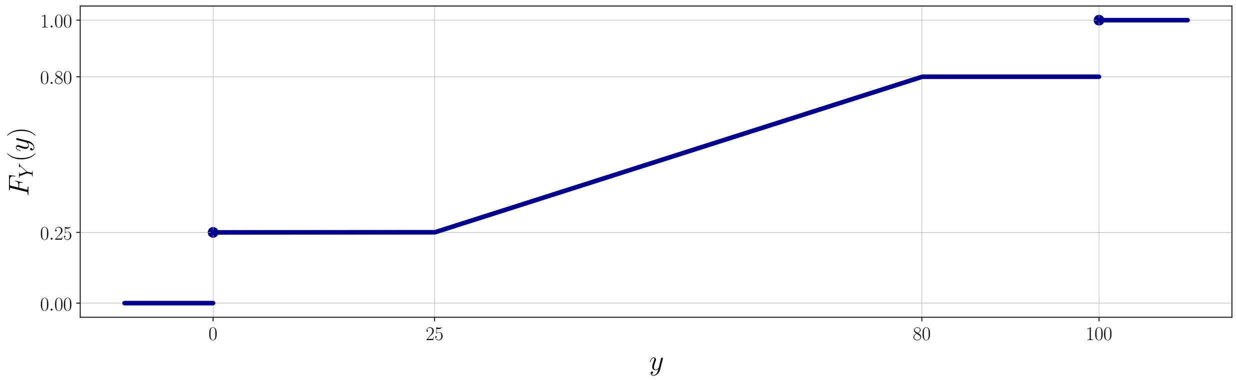


Figure 1: C.d.f. of the random variable Y.

2. Since $F_Y(y)$ has points of discontinuity $y = 0$ and $y = 100$, it only has generalized p.d.f. Taking derivative of c.d.f.:

$$f_Y(y) = \frac{dF_Y(y)}{dy} = \boxed{\frac{1}{100} \cdot I_{\{25 \leq y \leq 80\}} + \frac{1}{4}\delta(y) + \frac{1}{5}\delta(y - 100)},$$

where $\delta(y)$ is a Dirac delta function, defined as:

$$\delta(y) \simeq \begin{cases} +\infty, & y = 0, \\ 0, & y \neq 0, \end{cases} \quad \text{constrained by} \quad \int_{-\infty}^{+\infty} \delta(y)dy = 1.$$

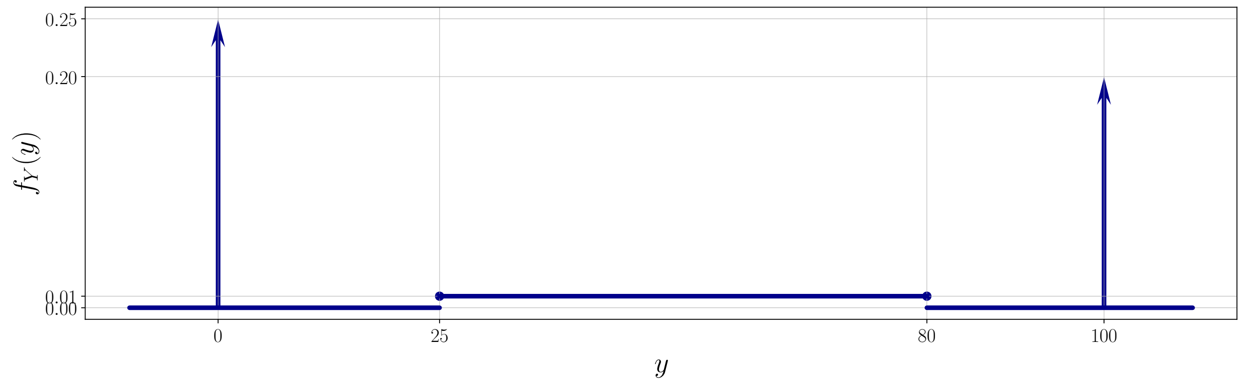


Figure 2: Generalized p.d.f. of the random variable Y.

A graph of $f_Y(y)$ is shown in the fig. 2.

The $f_Y(y)$ possesses the most important property of probability densities – normalization by 1:

$$\int_{-\infty}^{+\infty} f_Y(y)dy = \int_{25}^{80} \frac{1}{100}dy + \int_{-\infty}^{+\infty} \frac{1}{4}\delta(y)dy + \int_{-\infty}^{+\infty} \frac{1}{5}\delta(y-100)dy =$$

$$= \frac{y}{100}\Big|_{25}^{80} + \frac{1}{4}\int_{-\infty}^{+\infty} \delta(y)dy + \frac{1}{5}\underbrace{\int_{-\infty}^{+\infty} \delta(\eta)d\eta}_{\eta=y-100} = \frac{55}{100} + \frac{1}{4} + \frac{1}{5} = 1.$$

3. The mean and variance of a random variable $X \sim \mathcal{U}(a,b)$:

$$\mathsf{E}(X) = \frac{a+b}{2}, \qquad \mathsf{V}(X) = \frac{(b-a)^2}{12}.$$

In our case:

$$\mathsf{E}(X) = \frac{0+100}{2} = \boxed{50}, \qquad \mathsf{V}(X) = \frac{(b-a)^2}{12} = \frac{(100-0)^2}{12} = \boxed{\frac{2500}{3} \approx 833.333}.$$

There are 2 ways to calculate mean and variance of Y.

(I) **Direct calculation**

By definition:

$$\mathsf{E}(Y) = \int_{-\infty}^{+\infty} y f_Y(y)dy = \int_{25}^{80} \frac{y}{100}dy + \frac{1}{4}\int_{-\infty}^{+\infty} y\delta(y)dy + \frac{1}{5}\int_{-\infty}^{+\infty} y\delta(y-100)dy.$$

Using a sifting property of delta-function:

$$\int\limits_{-\infty}^{+\infty} \varphi(y)\delta(y-b)dy = \varphi(b),$$

we can proceed with calculation expected value of Y:

$$\mathsf{E}(Y) = \left.\frac{y^2}{200}\right|_{25}^{80} + \left.\frac{1}{4}y\right|_{y=0} + \left.\frac{1}{5}y\right|_{y=100} = \frac{231}{8} + 0 + 20 = \boxed{\frac{391}{8} = 48.875}.$$

In order to calculate variance $\mathsf{V}(Y)$, we need to find $\mathsf{E}\left(Y^2\right)$:

$$\begin{aligned}\mathsf{E}\left(Y^2\right) &= \int\limits_{-\infty}^{+\infty} y^2 f_Y(y)dy = \int\limits_{25}^{80} \frac{y^2}{100}dy + \frac{1}{4}\int\limits_{-\infty}^{+\infty} y^2\delta(y)dy + \frac{1}{5}\int\limits_{-\infty}^{+\infty} y^2\delta(y-100)dy = \\ &= \left.\frac{y^3}{300}\right|_{25}^{80} + \left.\frac{1}{4}y^2\right|_{y=0} + \left.\frac{1}{5}y^2\right|_{y=100} = \frac{19855}{12} + 0 + 2000 = \\ &= \frac{43855}{12} \approx 3654.583.\end{aligned}$$

Variance then:

$$\mathsf{V}(Y) = \mathsf{E}\left(Y^2\right) - \mathsf{E}(Y)^2 = \frac{43855}{12} - \left(\frac{391}{8}\right)^2 = \boxed{\frac{243037}{192} \approx 1265.818}.$$

(II) **Total expectation**

A random variable Y consists of 3 parts exhaustive, which is reflected in the total expectation equation:

$$\begin{aligned}\mathsf{E}(Y) &= \mathsf{E}(Y \mid X < 25)\cdot\mathsf{P}(X<25) + \mathsf{E}(Y \mid 25 \le X < 80)\cdot\mathsf{P}(25 \le X < 80) + \\ &+ \mathsf{E}(Y \mid X \ge 80)\cdot\mathsf{P}(X \ge 80) = 0\cdot\frac{1}{4} + \frac{25+80}{2}\cdot\frac{55}{100} + 100\cdot\frac{1}{5} = \\ &= 0 + \frac{231}{8} + 20 = \boxed{\frac{391}{8} = 48.875}.\end{aligned}$$

The same for $\mathsf{E}\left(Y^2\right)$:

$$\begin{aligned}\mathsf{E}\left(Y^2\right) &= \mathsf{E}\left(Y^2 \mid X < 25\right)\cdot\mathsf{P}(X<25) + \\ &+ \mathsf{E}\left(Y^2 \mid 25 \le X < 80\right)\cdot\mathsf{P}(25 \le X < 80) + \mathsf{E}\left(Y^2 \mid X \ge 80\right)\cdot\mathsf{P}(X \ge 80).\end{aligned}$$

While the first and the third terms are clear, the second one in the equation above requires more calculations:

$$\begin{aligned}\mathsf{E}\left(Y^2 \mid 25 \le X < 80\right) &= \mathsf{E}\left(X^2 \mid 25 \le X < 80\right) = \\ &= \mathsf{V}\left(X \mid 25 \le X < 80\right) + \mathsf{E}\left(X \mid 25 \le X < 80\right)^2 = \\ &= \frac{(80-25)^2}{12} + \left(\frac{25+80}{2}\right)^2 = \frac{9025}{3} \approx 3008.333.\end{aligned}$$

$\mathsf{E}\left(Y^2\right)$ then:

$$\mathsf{E}\left(Y^2\right) = 0 \cdot \frac{1}{4} + \frac{9025}{3} \cdot \frac{55}{100} + 100^2 \cdot \frac{1}{5} = \frac{43855}{12} \approx 3654.583.$$

Variance then:

$$\mathsf{V}(Y) = \mathsf{E}\left(Y^2\right) - \mathsf{E}(Y)^2 = \frac{43855}{12} - \left(\frac{391}{8}\right)^2 = \boxed{\frac{243037}{192} \approx 1265.818}.$$

4. Total expectation:

$$\mathsf{E}(Y) = \mathsf{E}(Y \mid Y > 0) \cdot \mathsf{P}(Y > 0) + \mathsf{E}(Y \mid Y \le 0) \cdot \mathsf{P}(Y \le 0).$$

Since $\mathsf{E}(Y \mid Y \le 0) = 0$:

$$\mathsf{E}(Y \mid Y > 0) = \frac{\mathsf{E}(Y)}{\mathsf{P}(Y > 0)} = \frac{391}{8} : \left(1 - \frac{1}{4}\right) = \boxed{\frac{391}{6} \approx 65.167}.$$

5. Using the definition of correlation coefficient:

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathsf{V}(X) \cdot \mathsf{V}(Y)}} = \frac{\mathsf{E}(XY) - \mathsf{E}(X) \cdot \mathsf{E}(Y)}{\sqrt{\mathsf{V}(X) \cdot \mathsf{V}(Y)}}.$$

The only term we don't know is $\mathsf{E}(XY)$. Since the joint distribution is unknown, let's use total expectation:

$$\begin{aligned}\mathsf{E}(XY) = {} & \mathsf{E}(XY \mid X < 25) \cdot \mathsf{P}(X < 25) + \mathsf{E}(XY \mid 25 \le X < 80) \cdot \mathsf{P}(25 \le X < 80) + \\ & + \mathsf{E}(XY \mid X \ge 80) \cdot \mathsf{P}(X \ge 80) = \mathsf{E}(0 \mid X < 25) \cdot \mathsf{P}(X < 25) + \\ & + \mathsf{E}\left(X^2 \mid 25 \le X < 80\right) \cdot \mathsf{P}(25 \le X < 80) + \mathsf{E}(100X \mid X \ge 80) \cdot \mathsf{P}(X \ge 80).\end{aligned}$$

From paragraph 3II we know that

$$\mathsf{E}\left(X^2 \mid 25 \le X < 80\right) = \frac{9025}{3} \approx 3008.333,$$

and by linearity of expected value:

$$\mathsf{E}(100X \mid X \geq 80) = 100\mathsf{E}(X \mid X \geq 80) = 100 \cdot \frac{80+100}{2} = 100 \cdot 90 = 9000.$$

Overall

$$\mathsf{E}(XY) = 0 \cdot \frac{1}{4} + \frac{9025}{3} \cdot \frac{55}{100} + 9000 \cdot \frac{1}{5} = \frac{41455}{12} \approx 3454.583.$$

Correlation coefficient then:

$$\mathrm{Corr}(X,Y) = \frac{\frac{41455}{12} - 50 \cdot \frac{391}{8}}{\sqrt{\frac{2500}{3} \cdot \frac{243037}{192}}} \approx \boxed{0.984}.$$

Problem 7

Let X and Y be two independent standard normal random variables. Find

1. $\mathsf{P}(|X+Y| > |X-Y|)$.
2. $\mathsf{P}(|X+Y| > 2\,|X-Y|)$.

Solution:

1. Using independence:

$$\begin{aligned}\mathsf{P}(|X+Y| > |X-Y|) &= \mathsf{P}\left((X+Y)^2 > (X-Y)^2\right) = \\ &= \mathsf{P}\left(X^2+2XY+Y^2 > X^2-2XY+Y^2\right) = \mathsf{P}(4XY>0) = \\ &= \mathsf{P}(X>0)\cdot\mathsf{P}(Y>0) + \mathsf{P}(X<0)\cdot\mathsf{P}(Y<0) = \\ &= \frac{1}{2}\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{2} = \boxed{\frac{1}{2}}.\end{aligned}$$

2. There are 2 ways to calculate $\mathsf{P}(|X+Y| > 2\,|X-Y|)$.

   (I) **Strict calculation**

$$\begin{aligned}\mathsf{P}(|X+Y| > 2\,|X-Y|) &= \mathsf{P}\left((X+Y)^2 > 4\,(X-Y)^2\right) = \\ &= \mathsf{P}\left((X+Y+2\,(X-Y))\cdot(X+Y-2\,(X-Y))>0\right) = \\ &= \mathsf{P}((3X-Y)\cdot(3Y-X)>0) = \\ &= \mathsf{P}(3X-Y>0\cap 3Y-X>0) + \\ &\quad + \mathsf{P}(3X-Y<0\cap 3Y-X<0).\end{aligned}$$

Random variables $3X-Y$ and $3Y-X$ are not independent, so the probabilities of intersections can not be separated. Let's consider a region $(3X-Y)\cdot(3Y-X)>0$. Its probability is a volume under a joint p.d.f. inside this region. This is illustrated in the fig. 3 and in the fig. 4.

Since X and Y are independent components of the vector $(X \quad Y)^\top \sim \mathcal{N}(\mathbf{0},\mathbf{I})$, their joint p.d.f. is radially symmetrical. The volume of a considered region is determined by the angle those lines cover.

The angle of one sector is $\arctan 3 - \arctan\frac{1}{3}$, so the required probability:

$$\mathsf{P}(|X+Y| > 2\,|X-Y|) = \frac{\arctan 3 - \arctan\frac{1}{3}}{\pi} = \boxed{\frac{\arctan\frac{4}{3}}{\pi} \approx 0.295}.$$

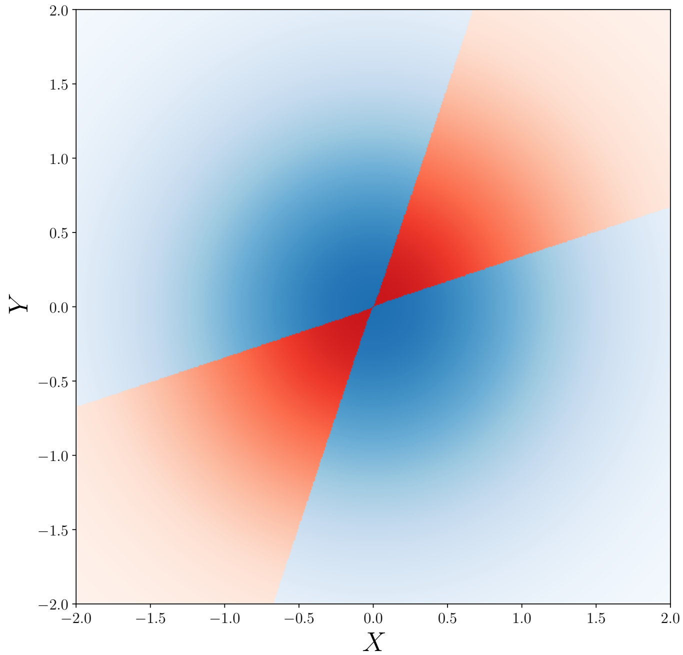


Figure 3: Probability density function of $(X \quad Y)^\top \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ (top view).

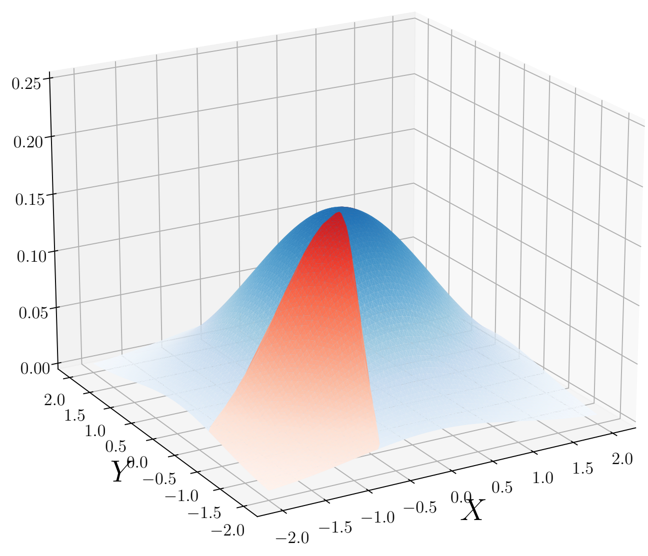


Figure 4: Probability density function of $(X \quad Y)^\top \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ (side view).

(II) **Approximation**

Let's consider variables $X+Y$ and $X-Y$. They have identical distribution:

$$X \pm Y \sim \mathcal{N}\left(0, 1^2 + (\pm 1)^2\right) = \mathcal{N}(0, 2).$$

Inherently, they are uncorrelated:

$$\begin{aligned}\operatorname{Cov}(X+Y, X-Y) &= \operatorname{Cov}(X,X) - \operatorname{Cov}(X,Y) + \operatorname{Cov}(Y,X) - \operatorname{Cov}(Y,Y) = \\ &= \mathsf{V}(X) - \mathsf{V}(Y) = 2 - 2 = 0.\end{aligned}$$

Since X and Y are components of the bivariate normal vector $(X \ \ Y)^\top \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, $X+Y$ and $X-Y$ are also components of a vector with bivariate normal distribution. In such case uncorrelatedness means independence – this is the unique property of multivariate normal distributions.

Thus, $X+Y$ and $X-Y$ are independent. The squares of standardized variables are also independent, and moreover, are χ^2-distributed:

$$Q_1 = \left(\frac{X+Y}{\sqrt{2}}\right)^2 \sim \chi^2_1, \qquad Q_2 = \left(\frac{X-Y}{\sqrt{2}}\right)^2 \sim \chi^2_1.$$

The initial probability can be rewritten with F-distribution:

$$\begin{aligned}\mathsf{P}(|X+Y| > 2\,|X-Y|) &= \mathsf{P}\left((X+Y)^2 > 4\,(X-Y)^2\right) = \mathsf{P}(Q_1 > 4Q_2) = \\ &= \mathsf{P}\left(\frac{Q_1/1}{Q_2/1} > 4\right) = \mathsf{P}(F_{1,1} > 4) \in \boxed{(\mathbf{0.25}, 0.5)}.\end{aligned}$$

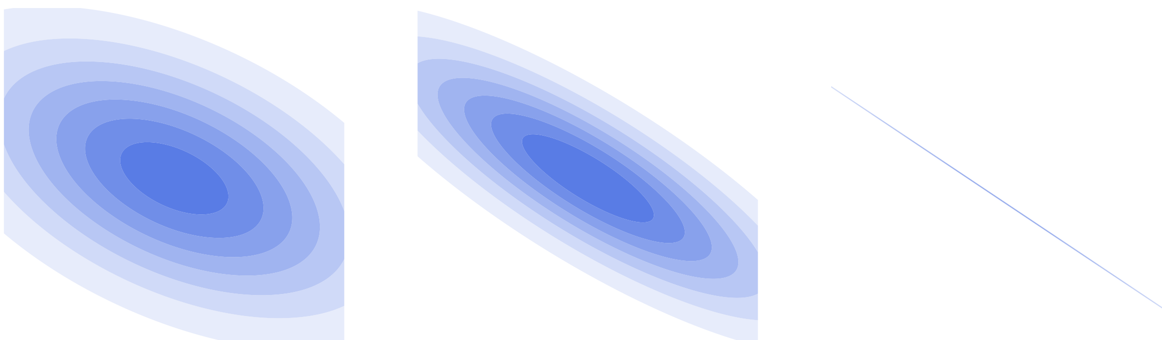
The exact result was found in 2I.

Problem 8

Two random variables are given: $X \sim \mathcal{N}(0, 9)$ and $Y \sim \mathcal{N}(0, 4)$. $\mathrm{Corr}(X, Y) = -1$.
Evaluate $\mathsf{P}(2X + Y > 3)$.

Solution:

Fig. 5 shows how a bivariate distribution changes with the increase of correlation coefficient between components.



(a) $\rho = -\frac{1}{2}$ (b) $\rho = -\frac{5}{6}$ (c) $\rho \to -1$

Figure 5: P.d.f. of a bivariate normal distribution with $\sigma_X = 3$ and $\sigma_Y = 2$.

The correlation of -1 means that there is a linear dependency between X and Y, but the change happens in opposite directions.
It means that standardized variables are opposite in sign:

$$\frac{X - \mu_X}{\sigma_X} = -\frac{Y - \mu_y}{\sigma_Y},$$

$$Y = \mu_Y - \frac{\sigma_Y}{\sigma_X}(X - \mu_X).$$

From problem statement:

$$Y = -\frac{2}{3}X.$$

Thus, the probability:

$$\mathsf{P}(2X + Y > 3) = \mathsf{P}\left(2X - \frac{2}{3}X > 3\right) = \mathsf{P}\left(\frac{4}{3}X > 3\right) = \mathsf{P}\left(X > \frac{9}{4}\right) =$$
$$= 1 - \Phi(2.25) \approx 1 - 0.988 = \boxed{0.012}.$$

Problem 9

The sample from bivariate normal distribution with random variables X and Y is following:

X	1.59	−2.20	−0.06	−1.45	−1.02	−2.59	−1.14	−3.25
Y	3.24	0.44	−1.14	5.40	2.09	5.33	1.25	8.72

Find 90% confidence interval for a population correlation coefficient ρ.

Solution:

$(1-\alpha)\cdot 100\%$ confidence interval for ρ:

$$\mathrm{CI}_{1-\alpha}(\rho) = \left(\tanh\left(\operatorname{artanh}(\widehat{\rho}) - z_{\alpha/2}\cdot\frac{1}{\sqrt{n-3}}\right); \tanh\left(\operatorname{artanh}(\widehat{\rho}) + z_{\alpha/2}\cdot\frac{1}{\sqrt{n-3}}\right)\right),$$

Values of sample mean:

$$\overline{x} = \frac{1}{8}\sum_{i=1}^{8} x_i \approx -1.265, \qquad \overline{y} = \frac{1}{8}\sum_{i=1}^{8} y_i \approx 3.166.$$

Values of corrected sums:

$$SS_{xx} = \sum_{i=1}^{8} x_i^2 - 8\cdot(-1.265)^2 \approx 16.283,$$
$$SS_{yy} = \sum_{i=1}^{8} y_i^2 - 8\cdot 3.166^2 \approx 71.340,$$
$$SS_{xy} = \sum_{i=1}^{8} x_i y_i - 8\cdot(-1.265)\cdot 3.166 \approx -17.240.$$

Value of sample correlation coefficient:

$$\widehat{\rho} = \frac{SS_{xy}}{\sqrt{SS_{xx}\cdot SS_{yy}}} = \frac{-17.240}{\sqrt{16.283\cdot 71.340}} \approx -0.506.$$

Value of Fisher-transformed correlation coefficient:

$$\operatorname{artanh}(\widehat{\rho}) = \frac{1}{2}\ln\left(\frac{1+\widehat{\rho}}{1-\widehat{\rho}}\right) = \frac{1}{2}\ln\left(\frac{1-0.506}{1+0.506}\right) \approx -0.557.$$

Confidence interval for Fisher-transformed ρ:

$$\begin{aligned}\mathrm{CI}_{90\%}(\operatorname{artanh}(\rho)) &= -0.557 \pm z_{0.05} \cdot \frac{1}{\sqrt{8-3}} = \\ &= -0.557 \pm 1.645 \cdot \frac{1}{\sqrt{8-3}} = \\ &= -0.557 \pm 0.736 = (-1.293; 0.179).\end{aligned}$$

Applying inverse Fisher-transform to the confidence interval above gives required interval:

$$\mathrm{CI}_{90\%}(\rho) = \tanh\left(\mathrm{CI}_{90\%}(\operatorname{artanh}(\rho))\right) = \frac{e^{2\mathrm{CI}_{90\%}(\operatorname{artanh}(\rho))} - 1}{e^{2\mathrm{CI}_{90\%}(\operatorname{artanh}(\rho))} + 1} =$$

$$= \left(\frac{e^{2\cdot(-1.293)} - 1}{e^{2\cdot(-1.293)} + 1}; \frac{e^{2\cdot 0.179} - 1}{e^{2\cdot 0.179} + 1}\right) = \boxed{(-0.860; 0.177)}.$$

Problem 10

Consider observations in the table below:

x	0	2	6	-3	4	1	-2	5	-1
y	8	2	0	6	1	5	7	3	4

1. Find Spearman's rank correlation coefficient r_s.
2. Find sample correlation coefficient r and compare it with r_s.

Solution:

1. Let's rank our sample and calculate differences d:

x	0	2	6	-3	4	1	-2	5	-1
y	8	2	0	6	1	5	7	3	4
rank(x)	4	6	9	1	7	5	2	8	3
rank(y)	9	3	1	7	2	6	8	4	5
d	-5	3	8	-6	5	-1	-6	4	-2
d^2	25	9	64	36	25	1	36	16	4

Spearman's rank correlation coefficient:

$$r_s = 1 - \frac{6\sum\limits_{i=1}^{n} d_i^2}{n\left(n^2-1\right)} = 1 - \frac{6 \cdot 216}{9 \cdot (81-1)} = \boxed{-0.8}.$$

2. Values of sample mean:

$$\overline{x} = \frac{1}{9}\sum_{i=1}^{9} x_i = \frac{4}{3}, \qquad \overline{y} = \frac{1}{9}\sum_{i=1}^{9} y_i = 4.$$

Values of corrected sums:

$$SS_{xx} = \sum_{i=1}^{9} x_i^2 - 9 \cdot \left(\frac{4}{3}\right)^2 = 80,$$

$$SS_{yy} = \sum_{i=1}^{9} y_i^2 - 9 \cdot 4^2 = 60,$$

$$SS_{xy} = \sum_{i=1}^{9} x_i y_i - 9 \cdot \frac{4}{3} \cdot 4 = -56.$$

Value of sample correlation coefficient:

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx} \cdot SS_{yy}}} = \frac{-56}{\sqrt{80 \cdot 60}} \approx \boxed{-0.808}.$$

Clearly, correlation coefficients are close due to the absence of prominent outliers:

$$\boxed{r \approx r_s}.$$