Topics covered in this tutorial

- Histogram density estimation
- Joint probability mass, density and cumulative distribution functions
- Marginal distributions and probabilities over regions
- Expectations of functions of two random variables
- Independence, covariance and correlation

Complete all questions in this tutorial before the tutorial test that follows. The test normally starts between approximately 16:00 and 16:15.

Onderwerpe wat in hierdie tutoriaal gedek word

- Histogramdigtheidskatting
- Gesamentlike waarskynlikheidsmassa-, digtheids- en kumulatiewe verdelingsfunksies
- Marginale verdelings en waarskynlikhede oor gebiede
- Verwagtings van funksies van twee toevalsveranderlikes
- Onafhanklikheid, kovariansie en korrelasie

Voltooi al die vrae in hierdie tutoriaal voor die tutoriaaltoets wat daarop volg. Die toets begin gewoonlik tussen ongeveer 16:00 en 16:15.

---

1. Two discrete random variables X and Y have the following joint probability mass function:

   Twee diskrete toevalsveranderlikes X en Y het die volgende gesamentlike waarskynlikheidsmassafunksie:

$P_{X,Y}(x,y)$	$y=0$	$y=1$	$y=2$
$x=1$	0.12	0.18	0.10
$x=2$	0.28	a	0.15
$x=3$	0.08	0.06	0.03

   (a) Find the value of a.

   Vind die waarde van a.

   (b) Calculate the marginal probability mass functions $P_X(x)$ and $P_Y(y)$.

   Bereken die marginale waarskynlikheidsmassafunksies $P_X(x)$ en $P_Y(y)$.

   (c) Are X and Y independent? Justify your answer.

   Is X en Y onafhanklik? Regverdig jou antwoord.

   (d) Calculate the joint CDF value $F_{X,Y}(2,1)$.

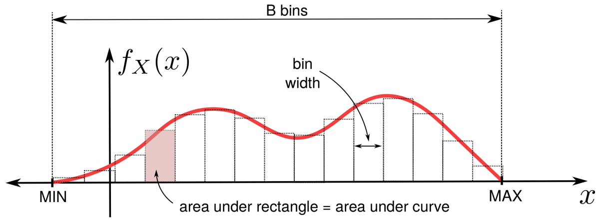
   Bereken die gesamentlike KVF-waarde $F_{X,Y}(2,1)$.

   (e) Calculate $P(X+Y \leq 3)$.

   Bereken $P(X+Y \leq 3)$.

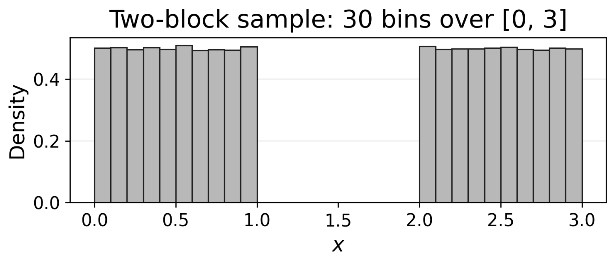
2. To construct a density histogram from N observations $\mathbf{x} = \{x_1, \ldots, x_N\}$, divide the data range into B equal-width, adjacent, non-overlapping bins. Count the observations in each bin and normalise each count by both N and the bin width. The resulting bar height estimates the probability density $f_X(x)$ over that bin, as illustrated below.

   Om 'n digtheidshistogram uit N waarnemings $\mathbf{x} = \{x_1, \ldots, x_N\}$ saam te stel, verdeel die databereik in B ewe wye, aangrensende, nie-oorvleuelende klasintervalle. Tel die waarnemings in elke interval en normaliseer elke telling met beide N en die intervalwydte. Die gevolglike staafhoogte beraam die waarskynlikheidsdigtheid $f_X(x)$ oor daardie interval, soos hieronder ge”illustreer.



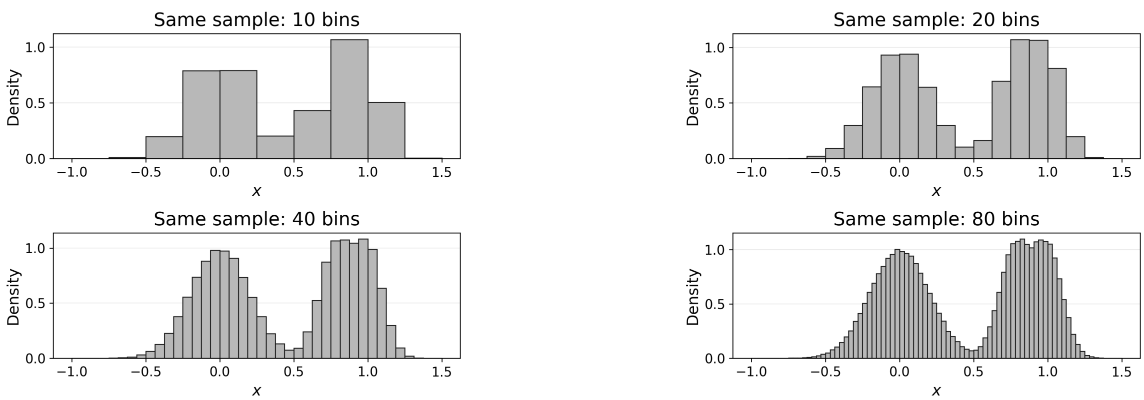
   (a) The histogram below was formed from a large sample of a random variable that is uniform on $[0, 1]$ with probability 0.5 and uniform on $[2, 3]$ with probability 0.5. The full interval $[0, 3]$ was divided into 30 equal-width bins. Explain why only 20 non-empty bars appear.

   Die histogram hieronder is gevorm uit 'n groot steekproef van 'n lukrake veranderlike wat met waarskynlikheid 0.5 eweredig oor $[0, 1]$ en met waarskynlikheid 0.5 eweredig oor $[2, 3]$ versprei is. Die volle interval $[0, 3]$ is in 30 ewe wye bakke verdeel. Verduidelik waarom slegs 20 nie-leë stawe verskyn.



   (b) Four histograms below were drawn from the same sample using 10, 20, 40, and 80 bins. For each panel, describe the effect of bin count on (i) smoothness and (ii) resolution of important shape features.

   Vier histogramme hieronder is uit dieselfde steekproef getrek met 10, 20, 40, en 80 bakke. Beskryf vir elkeen die effek op (i) gladheid en (ii) die vermoë om belangrike vormkenmerke uit te beeld.



   (c) Which bin count would you retain for presenting the shape, and why?

   Watter aantal bakke sou jy behou om die vorm aan te bied, en waarom?

   (d) A density histogram has $N = 200$ observations. The bin $[1.00, 1.25)$ has width $\Delta = 0.25$ and contains $n_j = 30$ observations. Because the vertical axis shows density rather than frequency, calculate the estimated bar height using $h_j = n_j/(N\Delta)$. No sketch is required.

   'n Digtheidshistogram het $N = 200$ waarnemings. Die klasinterval $[1.00, 1.25)$ het wydte $\Delta = 0.25$ en bevat $n_j = 30$ waarnemings. Omdat die vertikale as digtheid eerder as frekwensie toon, bereken die beraamde staafhoogte deur $h_j = n_j/(N\Delta)$ te gebruik. Geen skets word vereis nie.

3. Let X and Y be jointly continuous random variables with joint PDF

Laat X en Y gesamentlike kontinue toevalsveranderlikes wees met gesamentlike WDF

$$f_{X,Y}(x,y) = \begin{cases} cy + 1 & x, y \geq 0, x + y < 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Show the range of the random variables X and Y, S_{XY}, in the $x - y$ plane.

Toon die omvang van die ewekansige veranderlikes X en Y, S_{XY}, in die $x - y$-vlak.

(b) Find the constant c.

Vind die konstante c.

(c) Find the marginal PDFs $f_X(x)$ and $f_Y(y)$.

Vind die marginale WDFs $f_X(x)$ en $f_Y(y)$.

(d) Find $P(Y < 2X^2)$.

Vind $P(Y < 2X^2)$.

4. Let X and Y be continuous random variables with joint PDF:

Laat X en Y kontinue toevalsveranderlikes wees met gesamentlike WDF:

$$f_{X,Y}(x,y) = \begin{cases} 2 & 0 \leq y \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

(a) Calculate $\mathrm{E}[X]$, $\mathrm{E}[Y]$, and $r_{X,Y}$.

Bereken $\mathrm{E}[X]$, $\mathrm{E}[Y]$, en $r_{X,Y}$.

(b) Find $\mathrm{E}[2X + 3Y]$ and $\mathrm{E}[X^2 + Y^2]$.

Vind $\mathrm{E}[2X + 3Y]$ en $\mathrm{E}[X^2 + Y^2]$.

(c) Calculate the covariance $\mathrm{Cov}[X, Y]$ and correlation coefficient $\rho_{X,Y}$.

Bereken die kovariansie $\mathrm{Cov}[X, Y]$ en korrelasiekoëffisiënt $\rho_{X,Y}$.

(d) Use your covariance result to calculate $\mathrm{Var}[X + Y]$.

Gebruik jou kovariansieresultaat om $\mathrm{Var}[X + Y]$ te bereken.

(e) Define $U = 100X + 5$ and $V = -2Y + 7$. Without repeating any integrations, determine $\mathrm{Cov}[U, V]$ and $\rho_{U,V}$.

Definieer $U = 100X + 5$ en $V = -2Y + 7$. Bepaal $\mathrm{Cov}[U, V]$ en $\rho_{U,V}$ sonder om enige integrale te herhaal.

5. Let X and Y have the joint PDF

Laat X en Y die volgende gesamentlike WDF hê:

$$f_{X,Y}(x,y) = \begin{cases} x+y, & 0 \le x \le 1,\ 0 \le y \le 1, \\ 0, & \text{otherwise.} \end{cases}$$

(a) Derive the complete piecewise definition of the joint CDF $F_{X,Y}(x,y)$ for all real x and y.

Lei die volledige stuksgewyse definisie van die gesamentlike KVF $F_{X,Y}(x,y)$ vir alle reële x en y af.

(b) Find the marginal PDFs $f_X(x)$ and $f_Y(y)$.

Vind die marginale WDF'e $f_X(x)$ en $f_Y(y)$.

(c) Use the four-corner joint-CDF formula to calculate

$$P\left(0 < X \le \tfrac{1}{2},\ \tfrac{1}{4} < Y \le \tfrac{3}{4}\right).$$

Gebruik die vierhoekpuntformule vir die gesamentlike KVF om die volgende te bereken:

$$P\left(0 < X \le \tfrac{1}{2},\ \tfrac{1}{4} < Y \le \tfrac{3}{4}\right).$$

(d) Are X and Y independent? Justify your answer using the joint and marginal PDFs.

Is X en Y onafhanklik? Regverdig jou antwoord deur die gesamentlike en marginale WDF'e te gebruik.

6. Let X have a discrete uniform distribution with support $S_X = \{-1, 0, 1\}$, and define $Y = X^2$.

Laat X 'n diskrete eweredige verdeling met steunversameling $S_X = \{-1, 0, 1\}$ hê, en definieer $Y = X^2$.

(a) Determine the joint PMF of X and Y and present it as a table.

Bepaal die gesamentlike WMF van X en Y en stel dit as 'n tabel voor.

(b) Calculate $\mathrm{Cov}[X,Y]$ and hence state whether X and Y are uncorrelated.

Bereken $\mathrm{Cov}[X,Y]$ en meld gevolglik of X en Y ongekorreleerd is.

(c) Show by means of one decisive probability comparison that X and Y are nevertheless not independent.

Toon deur middel van een beslissende waarskynlikheidsvergelyking dat X en Y nogtans nie onafhanklik is nie.