Prof. Donna Ankerst, Katharina Selig

May 20th & 21st 2019

Problem 1

Let $X \sim \text{discrete}(X_1, X_2, \ldots, X_k)$ with $P(X = X_i) = \pi_i > 0$ for $i = 1, \ldots k$ with $\sum_{i=1}^{k} \pi_i = 1$.

(a) Write down the expressions for $\mathrm{E}[X]$ and $\mathrm{Var}(X)$.

(b) Show that

$$\mathrm{E}[aX + b] = a\mathrm{E}[X] + b,$$

for a, b scalar constants.

(c) Show that

$$\mathrm{Var}(aX + b) = a^2 \mathrm{Var}(X),$$

for a, b scalar constants.

Problem 2

Let $S = X_1 + X_2$, where $X_1 \sim \mathcal{N}(0, 1)$ and $X_2 \sim \mathcal{N}(4, 1)$ are independent.

(a) What is the distribution of $S = X_1 + X_2$?

`In R:`

(b) Generate a random sample of 1000 S's and graph a histogram of the sample. Comment on whether the distribution matches that in (a).

(c) Generate 1000 random samples of the mixture $f_X(x) = p f_{X_1}(x) + (1-p) f_{X_2}(x)$ for $p = 0.1, 0.4, 0.6, 0.8$ and graph the density estimate of the samples. For which values of p does the distribution of the mixture appear to be bi-modal?

Problem 3

Let $f_X = \sum_{j=1}^{6} \theta_j f_{X_j}$ be a mixture of Gamma distributions, where $X_j \sim \text{Gamma}(3, \lambda_j)$ are independent with rates $\lambda = (0.5, 1, 1.5, 2, 2.5, 3)$ and the following mixing probabilities $\theta = (0.1, 0.1, 0.2, 0.2, 0.3, 0.1)$.

(a) Calculate the mean and variance of $X \sim f_X$. You may use that the mean of a $\text{Gamma}(\alpha, \beta)$ random variable is $\frac{\alpha}{\beta}$ and the variance is $\frac{\alpha}{\beta^2}$.

`In R:`

(b) Write an efficient R code without loops to simulate 5000 observations from f_X.

(c) Create a plot including an empirical density estimate of the mixture (in red) as well as all component densities (in black).

Problem 4

We are given that for a random variable $X \in \mathbb{R}^p$ and $Y = CX + b$ for constants $C \in \mathbb{R}^{q\times p}$ and $b \in \mathbb{R}^q$ we have

- $\mathrm{E}[Y] = C\mathrm{E}[X] + b$
- $\mathrm{Var}(Y) = C\mathrm{Var}(X)C'$
- X normal distributed $\Rightarrow$ Y normal distributed.

(a) Suppose $Z \sim \mathcal{N}_p(0, I)$ for I the $p \times p$ identity matrix. Write down the pdf of Z.

(b) Let $Y \sim \mathcal{N}_p(\mu, \Sigma)$ with $\Sigma = Q'Q$ by Cholesky decomposition. Provide a linear transformation of Y from Z and show it has the correct mean and variance. Provide the dimensions of all quantities.

(c) Now suppose that it is of interest to efficiently generate n independent realizations of Y in a $n \times p$ data matrix X. Assume a $n \times p$ data matrix Z of independent $\mathcal{N}(0, 1)$ random variables is given. Write the transformation needed to convert Z to the data matrix X, proving that the rows have the correct distribution.

`In R:`

(d) Implement the algorithm from (c) in `R` for $n = 200$, $\mu = (1, 3, 0)$ and

$$\Sigma = \begin{pmatrix} 1.0 & -0.8 & -0.5 \\ -0.8 & 1.0 & 0.2 \\ -0.5 & 0.2 & 1.0 \end{pmatrix}$$

Use the `ggpairs` command from the package `GGally` to plot the data and check that covariances are approximately the same, reporting the max absolute differences. Repeat for $n = 1000$ and compare.

Problem 5

It has been stated that for $X \sim \mathrm{Poi}(\lambda)$ with $\lambda \sim \mathrm{Gamma}(r, \beta)$ we have $X \sim \mathrm{Negbin}(r, \frac{\beta}{\beta+1})$. Use the conditional mean and variance rules to derive the mean and variance of a $\mathrm{Negbin}(r, \frac{\beta}{\beta+1})$ distribution. You may use that the mean and variance are both λ for a $\mathrm{Poi}(\lambda)$ rv and $\frac{\alpha}{\beta}$ and $\frac{\alpha}{\beta^2}$ respectively for a $\mathrm{Gamma}(\alpha, \beta)$ rv.

Problem 6

Let X_1 and X_2 be independent exponentially distributed variables with parameter λ. Determine the distribution of $U = \dfrac{X_1}{X_1 + X_2}$ using the transformation rule. You may use that the density of $X \sim \mathrm{Exp}(\lambda)$ is $f(x) = \lambda e^{-\lambda x}$.