Solutions to the homework problems are to be left in the MAST30020 assignment box #200 on the ground floor in the Richard Berry Building (the boxes are located in the corridor leading to Wilson lab). Don’t forget to print your name, student ID, the subject name and code, and your lecturer’s name (K. Borovkov) on the first page of your solutions! All homework problems should be attempted. Only one (randomly chosen) of them will be marked. All material handed in must be on A4 size paper. Material on different sized paper will not be marked. The form and neatness of work can be considered in marking. Working and/or reasoning must be given to obtain full credit. The submission deadline is 5pm on Monday, 18 April 2016.

Tutorial Problems

1. Let $X$ be an integrable RV with median $m$. Show that $m = \arg\min_a E|X - a|$. 
   
   Hint: $|Y| = Y^+ + Y^-$; use the distribution tails to compute $EY^\pm$. Then minimise the resulting expression for $Y := X - a$ as a function of $a$. Fingers crossed, it will work.

2. Let $F$ be the DF of an RV $X \geq 0$. Show that, for any $\alpha > 0$,
   
   $EX^\alpha = \alpha \int_0^{\infty} x^{\alpha-1}(1 - F(x)) \, dx$.

   Hint: $Y := X^\alpha$ is a non-negative RV, too. How can one compute $EY$?

3. Let $X_1, X_2, X_3$ be i.i.d. exponential random variables with unit means. Compute the probability $P(X_1 \leq 2X_2 \leq 3X_3)$.

   Hint: If you don’t know what to begin with, try the total probability formula. It may happen that conditioning on some of the random variables will make computation feasible.

4. Let $X_1$ and $X_2$ be RVs such that $P(X_i = 1) = P(X_i = -1) = \frac{1}{4}$, $i = 1, 2$, and $P(X_1X_2 = 0) = 1$. Compute the expectations, variances and the covariance of $X_1$ and $X_2$. Are the $X_i$’s independent? Explain.

5. Let $X$ and $Y$ be two RVs with known means ($\mu_X := E X$, $\mu_Y := E Y$), variances ($\sigma_X^2 := \text{Var}(X)$, $\sigma_Y^2 := \text{Var}(Y)$) and correlation ($\rho := \text{Corr}(X,Y)$).

   (a) Derive the best (in mean quadratic) linear predictor for $Y$ from $X$, i.e. find the value $(\hat{a}, \hat{b}) := \arg\min_{(a,b)} E[Y - (aX + b)]^2$.

   (b) Compute the mean and the variance of the predictor $\hat{Y} := \hat{a}X + \hat{b}$.

   (c) Compute the covariance of $X$ and the prediction error $Y - \hat{Y}$. 

1
Homework Problems

1. For a positive RV $X$, show that:
   
   (a) $\mathbf{E}[(\ln X)^2] \geq [\ln(\mathbf{E}X)]^2$ for $X$ such that $^1 \mathbf{P}(X \leq e) = 1$;
   
   (b) $\mathbf{E}[(\ln X)^2] \leq [\ln(\mathbf{E}X)]^2$ for $X$ such that $\mathbf{P}(X \geq e) = 1$;
   
   (c) $\frac{\mathbf{E}Y}{\mathbf{E}X} \geq 1$.

2. Recall that an RV $X \sim \text{NB}(r, p)$ following the negative binomial distribution with parameters $r \in \mathbb{N}$ and $p \in (0, 1)$ has the following probability mass function: setting $q := 1 - p$,

   $\mathbf{P}(X = k) = \binom{k + r - 1}{k} q^r p^k, \quad k = 0, 1, 2, \ldots$

   Let $X_r \sim \text{NB}(r, p)$, $r = 1, 2, \ldots$, be independent RVs with negative binomial distributions with the respective parameters $r$ and a common $p \in (0, 1)$. Set $S := X_1 + X_2$. For a fixed $n \geq 0$, compute the conditional distribution $\mathbf{P}(X_2 = k \mid S = n)$, $k = 0, 1, 2, \ldots$, of $X_2$ given that $S = n$, $n = 0, 1, 2, \ldots$.

3. Let $(X, Y)$ be the coordinates of a point uniformly distributed in the region $D := \{(x, y) \in \mathbb{R}^2 : xy \leq 0, |x + y| \leq 2\}$.
   
   (a) Compute the density of $X$.
   
   (b) Compute the means $\mathbf{E}X$ and $\mathbf{E}Y$.
   
   (c) Compute the covariance matrix of $(X, Y)$.
   
   (d) How will your answers to parts (a)–(c) change if $(X, Y)$ is uniformly distributed over $D' := \{(x, y) \in \mathbb{R}^2 : xy \geq 0, |x + y| \leq 1\}$?
   
   (e) Compute the CEs $\mathbf{E}(Y \mid X)$, $\mathbf{E}(Y^2 \mid X)$, $\mathbf{E}(XY \mid X)$ and $\mathbf{E}(X^2 \mid X)$.

4. Let $X := (X_1, X_2)$ be a normal random vector with mean $(2, 1)$ and covariance matrix $C_X^2 = \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix}$. Using matrix algebra, find the distributions of (a) $Y = 3X_1 + 2X_2$ and (b) $Z = (Z_1, Z_2, Z_3)$ with $Z_1 = X_1 - X_2$, $Z_2 = X_1 + X_2$, $Z_3 = 3X_2$.

5. Let $X_1, X_2, \ldots$ be i.i.d. RVs with mean $\mathbf{E}X_j = \mu$. Put $S_n := X_1 + \cdots + X_n$, $n \geq 1$. Using the properties of conditional expectations, compute
   
   (a) $\mathbf{E}(S_n \mid X_1)$ and $\mathbf{E}(S_n \mid X_{n+1})$;
   
   (b) $\mathbf{E}(S_n \mid X_n, X_{n-1})$;
   
   (c) $\mathbf{E}(X_1 \mid S_n)$;
   
   (d) $\mathbf{E}(S_{n+m} \mid S_n)$, $m \geq 0$;
   
   (e) $\mathbf{E}(S_n \mid S_{n+m})$, $m \geq 0$.

Hints: (c) The observation that $\mathbf{E}(X_1 \mid S_n) = \mathbf{E}(X_2 \mid S_n) = \cdots = \mathbf{E}(X_n \mid S_n)$ (why is that so?) may prove to be useful. (e) You may wish to use the result of part (c).

^1Here $e$ is the base of the natural logarithm, a.k.a. Euler’s number, a.k.a. Napier’s constant.