KTH Matematik

Avd. Matematisk statistik

TENTAMEN I SF2940 SANNOLIKHETSTEORI/EXAM IN SF2940 PROBABILITY THEORY, Tuesday December 20, 2016, 08.00-13.00.

Examinator : Boualem Djehiche, tel. 08-7907875, email: boualem@kth.se

Tillåtna hjälpmedel/Permitted means of assistance: Appendix 2 in A. Gut: An Intermediate Course in Probability, Formulas for probability theory SF2940, L. Råde & B. Westergren: Mathematics Handbook for Science and Engineering and pocket calculator.

All used notation must be explained and defined. Reasoning and the calculations must be so detailed that they are easy to follow. Each problem yields max 10 p. You may apply results stated in a part of an exam question to another part of the exam question even if you have not solved the first part. 25 points will guarantee a passing result.

If you have received 5 bonus points from the home assignments, you may skip Problem 1(a). If you have received 10 bonus points, you may skip the whole Problem 1.

Solutions to the exam questions will be available at the course's homepage.

Good luck!

Problem 1

The random variables $N, X_1, X_2, \ldots$ are independent. N is Poisson distributed with mean $a/(1-p)$ and $X_1, X_2, \ldots$ are identically distributed with $P(X_1 = k) = p(1-p)^k$, $k = 0, 1, 2, \ldots$, where $0 < p < 1$, $a > 0$.

a) Determine the probability generating function of the sum $X_1 + \cdots + X_N$. (5 p)

b) Show that $X_1 + \cdots + X_N$ converges in distribution when $p \to 1$ and $a > 0$ fixed. Determine the limit distribution. (5 p)

Problem 2

Consider the pair of square integrable random variables (X, Y).

(a) Show that the conditional expectation $E[Y|X]$ minimizes the mean square error $E[(Y - h(X))^2]$ over all square integrable functions $h(X)$ of the random variable X. (4 p)

Assume that (X, Y) has the joint probability density, called Joakim's Valentine distribution which yields the plot ♡, given by

$$f_{X,Y}(x, y) = \frac{1}{2\pi} e^{-\frac{1}{2}\left(x^2 + \left(y - \sqrt{|x|}\right)^2\right)}, \quad x, y \in \mathbb{R}.$$

(b) Compute the conditional expectation $E[Y|X]$. (3 p)

(c) Compute the value of the optimal mean square error. (3 p)

Problem 3

Let X, Y, Z be independent and identically $N(0,1)$-distributed.

(a) Determine the distribution of each of the following random variables: $U = X + Y + Z$, $V = 2X - Y - Z$ and $W = Y - Z$. (3 p)

(b) Determine the distribution of the vector (U, V, W). (2 p)

(c) Determine $E[W^2U^2 + V^2U^3|U = 3]$. (2 p)

(e) Show that each of the random variables $X - Y, Y - Z$ and $Z - X$ is independent of U. (3 p)

Problem 4

For $n \geq 1$, let $\{X_k, 0 \leq k \leq n-1\}$ be a sequence of independent and identically distributed (i.i.d.) random variables such that $X_k \in N\left(0, \frac{1}{n}\right)$.

(a) Show that

$$\sum_{k=0}^{n-1} X_k^2 \xrightarrow{P} 1, \quad n \to \infty.$$

(5 p)

Set, for $\alpha > 0$,

$$S_n = \sum_{k=0}^{n-1} e^{\alpha \frac{k}{n}} X_k.$$

(b) Determine the distribution of S_n. (2 p)

(c) Show that S_n converges in distribution when $n \to \infty$. Determine the limit distribution. (3 p)

Problem 5

Let X have the $\Gamma(n, 1)$ distribution. $Y|X = x \in Po(x)$.

(a) Find the characteristic function of Y. (2 p)

(b) Find $E[Y]$ and $\mathrm{Var}(Y)$. (2 p)

(c) Show that $\frac{Y - E[Y]}{\sqrt{\mathrm{Var}(Y)}} \xrightarrow{d} N(0,1)$, when $n \to \infty$. (6 p)

KTH Matematik

Avd. Matematisk statistik

Suggested solutions to the exam Tuesday December 20, 2016, 08.00-13.00.

Problem 1

We have $N \in \mathrm{Po}(\frac{a}{1-p})$, $P(X_i = k) = p(1-p)^k$. Let $S_N := \sum_{i=1}^N X_i$.
Set $q := 1 - p$

a) We have $g_{S_N}(t) = g_N(g_X(t))$, where

$$g_X(t) = E[t^X] = \sum_{k=0}^{\infty} p(1-p)^k t^k = \frac{p}{1-qt}.$$

Hence,

$$g_{S_N}(t) = g_N\left(\frac{p}{1-qt}\right) = \exp\left(\frac{a}{q}\left(\frac{p}{1-qt} - 1\right)\right) = \exp\left(\frac{a(t-1)}{1-qt}\right).$$

(b) $g_{S_N}(t) \longrightarrow \exp a(t-1)$ as $p \to 1$, where $\exp a(t-1)$ is the probability generating function of $S \in \mathrm{Po}(a)$. Therefore,

$$S_N \xrightarrow{d} \mathrm{Po}(a), \quad p \to 1.$$

Problem 2

(a) We have

$$\begin{aligned} E[(Y - h(X))^2] &= E[(Y - E[Y|X] + E[Y|X] - h(X))^2] \\ &= E[(Y - E[Y|X])^2 + 2E[(Y - E[Y|X])\,(E[Y|X] - h(X))] + E[(E[Y|X] - h(X))^2] \\ &\geq E[(Y - E[Y|X])^2] + 2E[(Y - E[Y|X])\,(E[Y|X] - h(X))]. \end{aligned}$$

Now, since $E[Y|X] - h(X)$ is measurable w.r.t. the σ-algebra generated by X, we have

$$E[(Y - E[Y|X])\,(E[Y|X] - h(X))] = E[(E[Y|X] - h(X))\,E[(Y - E[Y|X])\,|X]] = 0.$$

Hence,

$$E[(Y - h(X))^2] \geq E[(Y - E[Y|X])^2]$$

for all h such that $h(X)$ is square-integrable. The optimal error is simply $E[(Y - E[Y|X])^2]$.

(b) We have $E[Y|X = x] = \int_{-\infty}^{+\infty} y f_{Y|X=x}(y)dy$ where

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(X, y)}{\int_{-\infty}^{+\infty} f_{X,Y}(x, z)dz}.$$

Note that

$$\int_{-\infty}^{+\infty} f_{X,Y}(x, z)dz = e^{-\frac{1}{2}x^2}\frac{1}{2\pi}\int_{-\infty}^{+\infty} e^{-\frac{1}{2}(z-\sqrt{|x|})^2}dz = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2},$$

since, $\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(z-\sqrt{|x|})^2}$ is the probability density of $N(\sqrt{|x|}, 1)$. Therefore,

$$f_{Y|X=x}(y) = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(y-\sqrt{|x|})^2}.$$

This means that $Y|X \in N(\sqrt{|X|}, 1)$ yielding the conditional mean $E[Y|X] = \sqrt{|X|}$ and variance $E[(Y - E[Y|X])^2] = 1$.

(c) The optimal mean error is $E[(Y - E[Y|X])^2] = 1$.

Problem 3

(a) and (b) Let $L := \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \in N(0, I), \;\; K := \begin{pmatrix} U \\ V \\ W \end{pmatrix}$ and $A := \begin{pmatrix} 1 & 1 & 1 \\ 2 & -1 & -1 \\ 0 & 1 & -1 \end{pmatrix}$. Then

$$K = AL \in N(0, AA^*),$$

where A^* is the transpose of the matrix A and $AA^* = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 2 \end{pmatrix}$. This means that the random variables U, V, W are independent and $U \in N(0, 3), \;\; V \in N(0, 6), \;\; W \in N(0, 2)$.

(c) By independence of the r.v. V, W and U we have $E[W^2U^2 + V^2U^3|U = 3] = 9E[W^2] + 27E[V^2] = 18 + 162 = 180$.

(d) $X - Y = \frac{V-W}{2}$, $Y - Z = W$ and $Z - X = -\frac{V+W}{2}$ are all functions of (V, W) which is independent of U.

Problem 4

(a) Note that $Y_k := \sqrt{n}X_k \in N(0, 1), \;\; k = 1, \cdots, n-1$ and are i.i.d. We have

$$\sum_{k=0}^{n-1} X_k^2 = \frac{1}{n}\sum_{k=0}^{n-1} Y_k^2.$$

By the Law of Large Numbers we obtain

$$\sum_{k=0}^{n-1} X_k^2 = \frac{1}{n}\sum_{k=0}^{n-1} Y_k^2 \xrightarrow{P} E[Y_1^2] = 1,$$

as $n \to \infty$.

(b) $S_n = \sum_{k=0}^{n-1} e^{\alpha\frac{k}{n}} X_k \in N\left(0, \frac{1}{n}\sum_{k=0}^{n-1} e^{2\alpha\frac{k}{n}}\right)$.

(c) Note that

$$\frac{1}{n}\sum_{k=0}^{n-1} e^{2\alpha\frac{k}{n}} = \sum_{k=0}^{n-1} e^{2\alpha\frac{k}{n}}\left(\frac{k}{n} - \frac{(k-1)}{n}\right) \longrightarrow \int_0^1 e^{2\alpha s}ds = \frac{1}{2\alpha}\left(e^{2\alpha} - 1\right), \quad n \to \infty.$$

Therefore, $S_n \xrightarrow{d} S \in N\left(0, \frac{1}{2\alpha}\left(e^{2\alpha} - 1\right)\right)$.

Problem 5

(a) We have $\varphi_Y(t) = E[e^{itY}] = E[E[e^{itY}|X]]$. But since $Y|X \in \mathrm{Po}(X)$, we obtain $E[e^{itY}|X] = e^{X(e^{it}-1)}$. Therefore,

$$\varphi_Y(t) = E[e^{X(e^{it}-1)}] = [is := e^{it} - 1,\ X \in \Gamma(n,1)] = (1 - is)^{-n} = (2 - e^{it})^{-n}.$$

We can also write

$$\varphi_Y(t) = (2 - e^{it})^{-n} = \frac{(1/2)^n}{(1 - \frac{e^{it}}{2})^n}$$

to conclude that $Y \in \mathrm{NBin}(n, \frac{1}{2})$.

b) We have $E[Y] = n$ and $Var(Y) = 2n$. They can be computed by differentiating the characteristic function at 0 or by just noting that $Y \in \mathrm{NBin}(n, \frac{1}{2})$.

c) The easiest way to prove that $\hat{Y} := \frac{Y - E[Y]}{\sqrt{\mathrm{Var}(Y)}} \xrightarrow{d} N(0,1)$, when $n \to \infty$ is to note that $Y \in \mathrm{NBin}(n, \frac{1}{2})$ can in fact be written as

$$Y = \sum_{i=1}^{n} X_i, \quad X_i's \quad \text{are i.i.d.}, X_i \in \mathrm{Ge}(\frac{1}{2}).$$

Then by the Central Limit Theorem we get the result. The other way is to compute the limit of the characteristic function. This goes as follows.

$$\begin{aligned}
\varphi_{\hat{Y}}(t) &= e^{\frac{-itn}{\sqrt{2n}}} \left(2 - e^{\frac{it}{\sqrt{2n}}}\right)^{-n} = \left[e^{\frac{it}{\sqrt{2n}}} \left(2 - e^{\frac{it}{\sqrt{2n}}}\right)\right]^{-n} \\
&= [x := it/\sqrt{2}] = \left(2e^{\frac{x}{\sqrt{n}}} - e^{2\frac{x}{\sqrt{n}}}\right)^{-n} = \left(2 + 2\tfrac{x}{\sqrt{n}} - 1 - 2\tfrac{x}{\sqrt{n}} - \tfrac{x^2}{n} + \circ(\tfrac{1}{n})\right)^{-n} \\
&= \left(1 - \tfrac{x^2}{n} + \circ(\tfrac{1}{n})\right)^{-n} \longrightarrow e^{x^2} = e^{-\frac{t^2}{2}},
\end{aligned}$$

showing that $\frac{Y - E[Y]}{\sqrt{\mathrm{Var}(Y)}} \xrightarrow{d} N(0,1)$.